DOCUMENT RESUME

ED 185 126 TM 800 353

AUTHOR Gorth, William Phillip; Perkins, Marcy R.
TITLE A Study of Minimum Competency Testing Programs. Final Program Development Resource Document.
INSTITUTION National Evaluation Systems, Inc., Amherst, Mass.
SPONS AGENCY National Inst. of Education (DHEW).
PUB DATE Dec 79
CONTRACT 400-79-0003
NOTE 184p.; For related documents, see TM 800 350-352.

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Academic Standards; Competence; Diffusion; Elementary Secondary Education; Instructional Design; Minimum Competencies; *Minimum Competency Testing; Program Administration; Program Effectiveness; Publicity; Test Construction; Testing Problems; *Testing Programs; Test Reliability; Test Selection; Test Validity; Validity

ABSTRACT

This resource document represents the integration of both practice and theory related to minimum competency testing (MCT), and is largely based on information collected in a nationwide survey of MCT programs. Chapter 1, To Implement or Not to Implement MCT, by Marcy R. Perkins, presents a definition of MCT and a discussion of the perceived benefits and costs of a MCT program. Chapter 2, Defining Competencies, by Perkins, presents the basic elements in the process of defining competencies, and describes how current programs are dealing with those issues. Chapter 3, Test Selection and Development, by Michael Priestley, discusses the initial decision to select or develop tests, and procedures in test selection and development, including the establishment of test reliability and validity. Chapter 4 Setting Standards, by Paula M. Nassif, describes standard setting strategies, including judgments on items and judgments on examinees. Chapter 5, Integrating Testing with Instruction, by Mary F. Tobin, discusses approaches to using test results for remedial, diagnostic purposes or curriculum development. Chapter 6, Program Management, by William Phillip Gorth and Peter E. Schriber, presents strategies for planning and personnel needs and costs of a MCT program. Chapter 7, Dissemination, by Schriber and Gorth, focuses on dissemination within and about a MCT program. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

NOTE TO THE READER

The materials contained in this report were prepared for the National Institute of Education (NIE), Department of Health, Education, and Welfare, under contract number (400-79-0003). This contract was awarded December 15, 1978, as the result of a competitive bidding procedure, to National Evaluation Systems, Inc. (NES), a firm that has developed and administered minimum competency tests under contract to State and local education agencies.

The purpose of this contract was to obtain previously unavailable descriptive information about minimum competency testing programs for the enlightenment of educators, researchers, and others interested in this area. Information on the consequences or impacts of these programs was not within the scope of work for this contract. However, NIE is currently planning a complementary study that will focus on program impacts.

In obtaining the descriptive information presented here, the NES project staff, during the spring of 1979, interviewed the directors of all State minimum competency testing programs and of 21 local district programs. Subsequent to these visits, NES staff developed written program descriptions, and these were sent to the program directors for verification. It is these verified program descriptions that form the basis for this report.

It should be emphasized that the information presented here provides a snapshot of the status of minimum competency testing programs as of June 30, 1979, and, owing to the dynamic nature of these programs, may not portray the programs as they are operating today.

Further, it should be emphasized that any opinions expressed in this report do not necessarily reflect NIE or HEW position or policy, and no endorsement of minimum competency testing or of any model described in this report by NIE or HEW should be inferred.

A Study of Minimum Competency Testing Programs

FINAL PROGRAM DEVELOPMENT

RESOURCE DOCUMENT

SUBMITTED BY:

William Phillip Gorth, Project Director
Marcy R. Perkins, Project Coordinator
National Evaluation Systems, Inc.
30 Gatehouse Road
Amherst, Massachusetts 01002

A PROJECT SPONSORED BY:

Office of Testing, Assessment and Evaluation
National Institute of Education
Dr. Judith S. Shoemaker, Project Officer

December 1979

TABLE OF CONTENTS

		Page
PREFACE		iv
OVERVIEW		1
CHAPTER 1:	TO IMPLEMENT OR NOT TO IMPLEMENT MCT Marcy R. Perkins	7
	Introduction What is Minimum Competency Testing? "To implement, or not to implement, MCT. . . ." Summary	
CHAPTER 2:	DEFINING COMPETENCIES Marcy R. Perkins	25
	Introduction Basic Elements in the Process of Defining Competencies Summary Guidelines for Defining Competencies: Three Examples	
CHAPTER 3:	TEST SELECTION AND DEVELOPMENT Michael Priestley	51
	Introduction Initial Decision: To Select or Develop Test Selection Test Development Establishing Validity and Reliability	

TABLE OF CONTENTS

		Page
CHAPTER 4:	SETTING STANDARDS Paula M. Nassif	93
	Introduction Issues and Parameters Standard Setting Strategies Judgments on Items Judgments on Examinees What is Actually Being Done	
CHAPTER 5:	INTEGRATING TESTING WITH INSTRUCTION Mary F. Tobin	124
	Introduction MCT Results and Decisions Related to Curriculum and Instruction Options for Organizing Instruction and Remediation Choosing the Appropriate Arrangements Integrating the Testing Program with Curriculum and Instruction	
CHAPTER 6:	PROGRAM MANAGEMENT William Phillip Gorth and Peter E. Schriber	139
	Introduction Personnel Costs	
CHAPTER 7:	DISSEMINATION Peter E. Schriber and William Phillip Gorth	153
	Introduction The Planning Process Documenting the Plan	

PREFACE

This resource document represents the integration of both practice and theory related to minimum competency testing. Information collected about the MCT programs that participated in this nationwide study served as much of the basis for descriptions herein about practices in use in the field, and the author(s) of each chapter brought a particular expertise to its content and structure. Every chapter was also submitted to a professional review by one or several of the other contributors to the document in order to ensure its accuracy, comprehensiveness, and usefulness to MCT program developers and reviewers.

While Sherry A. Rubinstein, Dolores R. Harris, and Richard Allan do not appear as authors in this document, they are to be acknowledged for their special contributions to it. Dr. Rubinstein, working with William Gorth, Marcy Perkins, and Mary Tobin, took a major leadership role in the conceptualization of both the content and structure of individual chapters, as well as to their integrity as one document. Dolores R. Harris accomplished the invaluable task of editing the chapters and in some cases contributed to a major restructuring of content. Dr. Allan contributed his expertise as a chapter reviewer. Finally, Mary Tobin is to be acknowledged here for her contributions to the document as a whole; she identified appropriate resources to be drawn upon in the document and worked with individual authors to construct chapter outlines.

C295

OVERVIEW

Marcy R. Perkins

Introduction

Because public concern about the condition of the American educational system has grown in recent years, more and more programs are being designed to assess whether students have acquired some specified set of skills to a predetermined minimum level. This trend toward minimum competency testing (MCT) has grown so fast, however, that educational decision makers are faced with the problem of designing and implementing such programs with little information as to what issues to consider, what questions to ask, and what decisions to make.

The major purpose of this resource document, therefore, is to provide information to help educational decision makers on all levels make informed choices about minimum competency testing. The document is designed to present a range of options that have been tried in the field, and to present issues that have arisen in the course of implementing MCT programs. It can serve as a resource for discussions about minimum competency testing or for its implementation.

The document is likely to be most useful to those for whom a decision has been made, on whatever level, to develop and implement a minimum competency testing program or to review the adequacy of a proposed or existing program. The document is intended for a wide range of audiences, from state legislators to state education department staff to local district administrators, teachers, and consultants. The goal is to reach anyone who has an interest in, or is responsible for any part of a minimum competency testing program.

Implementation Issues to Consider

Regardless of the purpose or level of involvement an educator has with respect to a minimum competency testing program, a thorough consideration of such questions as the following may help in making whatever decisions there are to be made:

-- What kinds of competencies shall we define (e.g., life skills, basic skills)?

-- Who will have responsibility for defining the competencies?

-- How do we set standards?

-- What standards shall we set?

-- Do we develop or select tests? How do we do either?

-- If we develop a test, how do we ensure its fairness?

-- Shall we have different tests/standards/competencies for racial groups/ethnic groups/special education students/limited English-speaking students?

-- Who is to administer the tests?

-- What kinds of scores do we want to compute?

-- Who do we report results to?

-- Do we disseminate just test results, or the tests themselves? How does this decision affect test development?

-- How do we use what money we have most effectively?

-- What is a good way to manage this program?

-- Do we want to build in formative/summative evaluation of the program? Shall we systematically study the impacts of our program?

-- How will we know if and when our goals have been met?

-- After MCT, what?

While it might be worthwhile to treat all of the issues in detail, and so satisfy all the needs of any program developer, the resulting encyclopedic document might no longer be timely, and might also be so weighty and unwieldy as to function only as a 100-pound bookend on program developers' shelves. The topics discussed in this document, therefore, which are only a subset of those which could be discussed, were selected on the basis of a needs analysis conducted during site visits to more than 50 MCT programs and on the basis of the needs which program developers expressed at national conferences.

General Chapter Characteristics

This document is intended to be nonevaluative, and therefore no single perspective will be advocated on any issue. Rather, the salient issues related to MCT that have been identified through the site visits to operating programs are described. Since the document is also intended to be practical, instead of strictly academic, it will present examples of procedures and materials used by local and state agencies to help illustrate what can be done to resolve the issues under discussion. It is important to note, however, that in cases where specific practices are mentioned or materials cited, these references are not in any way endorsements of the particular procedures or documents. Finally, the authors herein do not assume that all readers are always familiar with the terminology of educators and measurement specialists. Therefore, to avoid confusion or ambiguity, technical terms or terms with very specific usages are also defined in the context of the particular chapter in which they occur.

Document Framework

Discussions about minimum competency testing programs generally revolve around the various components of these programs and the activities associated with developing these components. In this document, while more components and activities are discussed than may be reflected in chapter titles, not all possible components or activities are included because of

space limitations. In order to help the reader access information of interest, a number of components generally associated with minimum competency testing programs are listed below. Next to each is the chapter number in which some discussion of that component can be found.

Components	Chapter
Policy	-
Program Purposes	1
Competencies	2
Measurement Instruments	3
Standards	4
Target Groups	2
Testing Schedule	6
Test Administration	6
Scoring and Analysis	6
Reporting and Dissemination	7
Use of Data	5
Testing Special Populations	-
Remediation	5
Program Staffing and Management	6
Strategies for Cost Effectiveness	6
Program Evaluation	-

Summary of Remaining Chapters

Chapter 1: To Implement or not to Implement MCT

The major intents of this first chapter are to provide a definition of MCT that will serve as the basis for the remaining chapters, and to present the myriad of issues that have arisen in the field about whether or not MCT should be implemented on any level. The perceived costs and benefits of MCT that have been expressed by program personnel, testing specialists, and the public are discussed.

Chapter 2: Defining Competencies

The purpose of this chapter is to present issues related to the definition of competencies and to describe how programs in the field are currently dealing with the issues. Discussed are a number of questions programs are considering that concern the orientation of competencies, who may be involved in the identification process, and how validation may take place.

Chapter 3: Test Selection and Development

The primary purpose of this chapter is to present issues being faced by programs that are related to making a decision to either select or develop test instruments. Also discussed are the issues related to implementing either option.

Chapter 4: Setting Standards

The aim of this chapter is to describe standard setting strategies used in the field and to present issues concerning the selection of one or another strategy.

Chapter 5: Integrating Testing with Instruction

Since a frequently expressed goal of minimum competency testing is to identify students who need remediation, Chapter 5 discusses approaches to using test results for remedial, diagnostic purposes. It also deals with the integration of test results with instruction and the development of instruction.

Chapter 6: Program Management

The major purpose of this chapter is to present issues related to the management of a minimum competency testing program, either at the state or local level. A discussion of cost effectiveness strategies is also included.

Chapter 7: Dissemination

The last chapter focuses on issues related to dissemination within and about a minimum competency testing program, and also considers the question of how those directly affected by a program can be kept informed of its activities and how the program can be presented to the public.

CHAPTER 1

TO IMPLEMENT OR NOT TO IMPLEMENT MCT

Marcy R. Perkins

Introduction

As mentioned in the Overview, minimum competency testing is a fast-growing educational phenomenon that continues to spread even in the face of little information as to how programs may be developed and implemented or what effects they may be having. While this entire document is intended to help bridge that informational gap by presenting some of the issues being faced in the field and discussing the ways in which programs are resolving them, this chapter serves two specific purposes as a preliminary to the other chapters.

First, since minimum competency testing "means many things to many people" (Airasian, Pedulla, & Madaus, 1978), one intent of this chapter is to provide a working definition of MCT. This definition, only one of the many formulations possible, is based on the features observed and accepted in the field which served as the basis for selecting the programs in the study. Second, since this document is intended for all policy makers, not just for those who have already implemented competency testing, issues related to the question of whether or not MCT should be implemented will be discussed in this chapter. Before turning to these, however, a number of general points need to be discussed.

It is assumed here that systematic attempts to consider the issues, both for and against the implementation of minimum competency testing, will result in sounder decisions. This does not mean, however, that decision makers in states and local districts which have already adopted such programs cannot benefit from the material presented in this chapter. The issues discussed may serve to shed light on both unresolved issues and implementation difficulties that result from the failure of a program to deal with the reservations of key individuals or groups.

Because of the necessary limitations of space, this chapter does not discuss every one of the issues related to the perceived costs and benefits of minimum competency testing. Moreover, no single perspective

will be advocated with respect to any of the issues raised, nor will a stance be taken on the issue of whether to implement or not to implement MCT. Finally, those interested in the history of minimum competency testing are destined to be disappointed if they search for it here. An account of the background and development of MCT is not likely to be as helpful for program developers as a systematic presentation and discussion of the strengths and weaknesses of MCT as seen by those in the field.

What is Minimum Competency Testing?

> "When I use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean--neither more nor less."
>
> "The question is," said Alice, "whether you *can* make words mean so many different things."
>
> "The question is," said Humpty Dumpty, "which is to be master--that's all" (Lewis Carroll, Through the Looking Glass).

If there is one point upon which all testing specialists, program administrators, and educational policy makers agree, it is that there is no consistent terminology for minimum competency testing in use in the testing field. "Standards" in some programs can mean "competencies" in others; "competencies" themselves can be synonymous with "competency areas," "objectives," "skill statements," and "performance indicators," to cite only a few terms among many. With this wealth of terminology, some of which is specific to only a few programs, how then is minimum competency testing defined? Are there components which are common to all programs?

Table 1 presents the texts of nine definitions of MCT found in the research and policy literature. In the first five, there is a clear emphasis on student acquisition of certain minimum skills, and on assessment of that achievement. In the sixth and seventh definitions, potential effects of minimum competency testing, rather than its strict defining characteristics, are delineated. In the last two, the specific components and procedures of minimum competency testing programs are presented. Even in these, however, the concept of some kind of a standard is evident.

TABLE 1

Definitions of Minimum Competency Testing
Employed in the Field

- Minimum competency testing programs are "organized efforts to make sure public school students are able to demonstrate their mastery of certain minimum skills needed to perform tasks they will routinely confront in adult life."

(AFSC, 1978)

- Minimum competency tests are constructed to measure the acquisition of competence or skills to or beyond a certain defined standard.

(Miller, 1978)

- Minimum competency testing programs are "testing programs which attempt to learn whether each student is at least 'minimally competent' by the time the student graduates from public school."

(NSBA, 1978)

- Minimum competency testing is "a certification mechanism whereby a pupil must demonstrate that he/she has mastered certain minimal (sic) skills in order to receive a high school diploma."

(Airasian et al., 1978)

- Minimum competency testing is "a device to increase emphasis on the three R's or basics."

(Airasian et al., 1978)

TABLE 1 (continued)

- Minimum competency testing is "a mechanism for tightening up promotion requirements; certifying early exit from the school system; holding educators responsible for poor student achievement; increasing the cost-effectiveness of education; identifying and remediating pupils who have learning difficulties; or increasing the public's confidence in the schools and their graduates."

(Airasian et al., 1978)

- Nearly all minimum competency testing programs seek "to define minimum learning outcomes for students in a variety of academic areas" and "to insure that these standards are satisfied."

(Cohen & Haney, 1978)

- Minimum competency testing involves:

  (1) the use of objective, criterion-referenced competency tests;

  (2) the assessment of reading and computation using "real life" or "life skill" items;

  (3) the requirement of a specified mastery level for high school graduation;

  (4) the early introduction of such testing for purposes of identification and remediation.

(Elford, 1977)

- Competency-based education (used in this paper nearly synonymously with minimum competency testing) is "a data-based, adaptive, performance-oriented set of integrated processes that facilitate, measure, record, and certify within the context of flexible time parameters the demonstration of known, explicitly stated, and agreed upon learning outcomes that reflect functioning in life roles."

(Spady, 1977)

For the purposes of the NIE study of minimum competency testing programs, two features were selected as being distinctive of MCT programs. Programs can, and do, vary widely on a great number of dimensions, but to be included in the study, any program under consideration had to have at least the following two features:

(1) the presence of an explicit standard for determining acceptable performance; and

(2) the use of test results to make decisions about individual students.

No other features were taken into account, such as the reasons for initiating a program (e.g., certification of students for graduation, grade promotion decisions, identification of students in need of remediation), or the grade levels set for testing (e.g., high school grades only; a mix of elementary, junior high and high school grades, elementary grades only).

The presence of a standard gives meaning to the concept of pass/fail, and so distinguishes MCT from statewide assessments. In the latter, student achievement may be monitored individually (although many assessments use sampling rather than census testing), but not with respect to any specific standard; i.e., a student does not pass or fail the tests. Student results are generally reported according to groups if sampling is used, rather than by individuals. If individual results are reported, they are usually interpreted at the discretion of individual teachers. In minimum competency testing, by contrast, students are required to achieve certain minimum standards of performance; that there are specific consequences to students for meeting or not meeting the standards is the second distinctive feature of MCT.

In the programs of the study, consequences to students who achieve the minimum standards may range from the receipt of a high school diploma or certificate of special recognition to promotion from grade to grade. Consequences for not meeting the standards can include compulsory enrollment in remedial classes, grade retention, or the receipt of a certificate of school attendance instead of a high school diploma. Regardless of the importance of the consequences or whether they are applied for passing vs. failing the tests, the fact remains that some kind of consequences are present in programs accepted as minimum competency testing programs.

"To implement, or not to implement, MCT. . . ."

Minimum competency testing is, without question, one of the most hotly debated subjects in the world of testing today. Proponents make strong claims about its potential benefits, and opponents argue just as strongly about its potentially harmful effects. It is not the purpose of this chapter to determine, once and for all, the various impacts of MCT or whether they are harmful or beneficial. Rather, the intent is to present major issues for policy makers to consider as they make decisions about whether MCT will serve the particular goals and purposes established for their testing programs. For policy makers on the point of making a decision about minimum competency testing, weighing the advantages and disadvantages of MCT, especially as these relate to a particular program, will help to reach decisions that are well-informed and reasoned. One of the chief criticisms of MCT programs today concerns the speed with which implementation has been required, a speed which has not always allowed program developers the time to plan as carefully as they might like.

Because this chapter is to be nonevaluative and impartial in its discussions of the issues, it is hard to know which side of the controversy to present first. Beginning with either the pro- or the anti-MCT arguments could be construed as presenting, however subtly, a specific position on the issues. Therefore, a decision was made to determine the order of presentation by flipping a coin: heads, the pro-MCT arguments go first; tails, the anti-MCT arguments go first. The coin turned up heads.

Perceived Benefits of Minimum Competency Testing

Listed in Table 2 are a number of perceived benefits of minimum competency testing that have been culled from a wide variety of sources, including the research literature, MCT program publications, professional conference proceedings, and personal communications during the site visits conducted in this study. Each of these has been cited as a benefit or potential purpose or useful effect of minimum competency testing by at least one person in the field. Most have been cited any number of times as reasons for implementing MCT either locally or statewide. The benefits appear to fall into a finite set of types: MCT may (1) restore confidence in the high school diploma, (2) involve the public in education, (3) improve teaching and learning, (4) serve a diagnostic, remedial function, and (5) provide a mechanism of accountability.

TABLE 2

Perceived Benefits of Minimum Competency Testing

- restores meaning to a high school diploma
- reestablishes public confidence in the schools
- impels us to face squarely the question of "what is a high school education?"
- sets meaningful standards for diploma award and grade promotion
- challenges the validity of using seat time and course credits as basis for certifying student accomplishments
- certifies that students have specific minimum competencies
- involves the public and local educators in defining educational standards and goals
- focuses the resources of a school district on a clear set of goals
- defines more precisely what skills must be taught and learned—for students, parents, and teachers
- promotes carefully organized teaching and carefully designed sequential learning
- reemphasizes basic skills instruction
- helps promote competencies of life after school
- broadens educational alternatives and options

TABLE 2 (continued)

- motivates students to master basic reading, mathematics, and writing skills
- stimulates teachers and students to put forth their best efforts
- identifies students lacking basic skills at an early stage
- encourages revision of courses to correct identified skill deficiences
- ensures that schools help those students who have the greatest educational need
- can bring about cohesiveness in teacher training
- can truly individualize instruction
- shifts priorities from process to product
- holds schools accountable for educational products
- furnishes information to the public about performance of educational institutions
- provides an opportunity to remedy the effects of discrimination by identifying learning problems early in the educational process
- provides greater holding power for students in the senior year
- provides for easier allocation of resources

Let us consider first the view that minimum competency testing can restore confidence in the high school diploma. It has been apparent for some time that there is widespread public disillusion and dissatisfaction with the quality of American education. Employers complain that applicants with high school degrees are unable to complete job applications correctly. Colleges and universities complain that they must institute remedial reading classes in order to raise the reading ability of incoming students to levels high enough for college work. The public points to declining test scores as an indication of the inadequate skills which students possess at graduation. In the light of this evidence, all segments of the public are concerned to know what a high school diploma actually certifies about the skills of the student. And MCT is seen as a way of clearly and precisely demonstrating what students can do and of ensuring that they have those "minimum" skills necessary to function in society (e.g., AFSC, 1978; NSBA, 1978). An auxiliary benefit is that along with a precise definition of skills and a demonstration that students indeed have those skills will come a greater public confidence in the educational system (e.g., AFSC, 1978; Nickse, 1978).

According to Walker (1978), the main support for MCT has come from the public, and the second category of perceived benefits relates to the involvement of the public in educational goal setting. Proponents of MCT cite as one of its benefits the fact that responsibility for defining the goals and intended outcomes of a high school education is shared by educators and the public (e.g., NSBA, 1978; Nickse, 1978). It is certainly the case that, in most MCT programs, administrators have considered it important to involve representatives from such constituencies as parents, the business community, and outside educational organizations. Frequently, surveys of these groups have also been conducted for the purpose of providing input to the processes of defining and/or validating competencies and setting standards.

The realms of teaching and learning comprise a third area in which its proponents consider that minimum competency testing will have a beneficial impact. Since a legal question may arise as to whether one may test a skill that has not been directly taught, many supporters see MCT as an impetus to a careful examination of the curriculum in light of the goals of the MCT program (e.g., AFSC, 1978). Other MCT advocates believe that a reemphasis of the basic skills is in order and can be accomplished through minimum competency testing (e.g., NSBA, 1978). Still others, who advocate a systems or competency-based approach to education, consider MCT to be the means for restructuring curricula to reflect such an approach. Finally, there are those who feel that MCT will increase the motivational levels of both students and teachers (e.g., NSBA, 1978).

Related to the hope that MCT will help to improve teaching and the curricula is the expectation that it will stimulate the establishment of remedial programs for students shown to be deficient in the basic skills (e.g., NSBA, 1978; AFSC, 1978; Wilson, 1976). In many MCT programs, the major goal of testing is to identify those students who need additional instruction; the intended remedy for deficiency is most often remediation.

Finally, although some MCT programs specifically forbid the use of test results for accountability purposes, accountability is still a live issue in the field of education, and MCT is seen as one way of establishing accountability. Students, teachers, and administrators alike can be held accountable for their respective educational responsibilities (e.g., Scott, 1978).

Perceived Costs of Minimum Competency Testing

Enumerated in Table 3 are the perceived disadvantages of MCT which are commonly cited by opponents of minimum competency testing. Like the perceived benefits, the perceived costs center on the potential effects of MCT on a variety of elements, and these effects are seen to be harmful in some way. Once again, the discussion may be facilitated by grouping the points according to the element affected. Therefore, perceived disadvantages may be seen in terms of the potential harmful effects of MCT on (1) various populations of students, (2) the curriculum, (3) teachers and administrators, and (4) control of education.

With respect to its effects on various student populations, the criticisms of minimum competency testing are several. Opponents of MCT believe that underachievers, diagnosed as being "below competency standards," will suffer from further labeling, especially if the receipt of a standard high school diploma is contingent upon passing a competency test. On the other hand, it is claimed that average students are unrecognized and gifted students go unchallenged in MCT programs (AFSC, 1978). Advocates of racial, ethnic, or special education students assert that competency testing may promote bias against these groups, especially if school systems are believed to be already segregated or discriminatory against these student populations in some other way (Airasian et al., 1978; Scott, 1978). Finally, minimum competency testing may unfairly place the burden of failure squarely on the student, rather than making failure a shared responsibility of student, teacher, and school system (AFSC, 1978).

TABLE 3

Perceived Costs of Minimum Competency Testing

- emphasis on the practical will lead to an erosion of liberal education
- causes less attention to be paid to difficult-to-measure learning outcomes
- promotes teaching to the test
- will be the "deathknell for the inquiry approach to education"
- oversimplifies issues of defining competencies and standards and of granting credentials to students
- promotes confusion as to the meaning of the high school diploma when competency definition is left to local districts
- fails to adequately consider community disagreement over the nature and difficulty of competencies
- will exclude more children from schools and further stigmatize underachievers
- will cause "minimums" to become "maximums," thus failing to provide enough instructional challenge in school
- may unfairly label students and cause more of the "less able" to be retained
- may cause an increase in dropouts, depending on the minimum that is set
- provides no recognition of the "average" student

TABLE 3 (continued)

- fails to provide alternatives that can "inspire" average students to excel in some areas
- ignores the special needs of gifted students, giving them less opportunity to be challenged and to expand their horizons
- may have adverse impact on a student's future career as a result of a withheld diploma
- may promote bias against racial, ethnic, and/or special needs groups
- places the burden of "failure" on the student
- causes educators to be held unfairly accountable
- intensifies the conflict for educators between humaneness and accountability
- increases the record keeping burden for administrators
- does not assure that students will receive effective remediation
- does not assure that all of the perceived needs and benefits will be met and realized
- promotes the power of the state at the expense of local district autonomy
- can be costly, especially where implementation and remediation are concerned

"Minimums will become maximums!" is a commonly expressed fear about the effect of minimum competency testing on curriculum. Most educators admit that it is difficult to define "*minimum* competency," and therefore, critics raise questions about what a diploma can really mean if different definitions of competency are derived by individual local districts (NSBA, 1978). There are also fears that MCT may lead to a narrow and overly limited curriculum, because of the emphasis which such programs seem to place upon a certain few basic skills and upon those skills which lend themselves to definition in measurable terms.

Issues of teacher and school accountability seen by some as beneficial are seen by others as harmful effects of minimum competency testing. Opponents of MCT assert that educators are often held unfairly accountable and that minimum competency testing only serves to intensify the conflict between "humaneness" and "accountability" in the role of the educator (ASCD, 1978; NSBA, 1978). Furthermore, the initiation of MCT may unfairly place additional burdens upon school teachers and administrators in the form of extra record keeping and, in some cases, mandatory curriculum reform (NSBA, 1978). Already busy school personnel, in other words, will be expected to assume additional roles and tasks with the effect, perhaps, of decreasing their time to produce enriched curricula.

Finally, its effect on control of education is seen to be a disadvantage of minimum competency testing (Nickse, 1978). In many states, local autonomy is a valued prerogative, and MCT mandated on the state level is seen as an infringement of that prerogative. Local school districts also complain that the states often impose certain requirements and yet give little or no financial or technical support to help the local districts comply. This same argument can also apply at the state level, since in some cases the legislature may enact certain requirements and yet fail to appropriate funds to support compliance.

Criticism of a Different Nature

In addition to the perceived advantages and disadvantages of minimum competency testing enumerated above, writers in the field have offered other criticisms of a somewhat different nature. Those costs and benefits already discussed are generally predicated on the assumptions (1) that the identification and definition of competencies and minimum standards of performance is a straightforward process, and (2) that principles and techniques exist for the construction of reliable and valid test instruments.

The other criticisms, by contrast, tend to focus on these two assumptions and also on the actual implementation procedures for minimum competency testing.

With respect to the first assumption, some critics have taken issue with the "seductive nature" of the vocabulary used in minimum competency testing programs. "Undefined, perhaps undefinable terms are used without consideration in discussing MCT programs, and it is only when one thinks through the meaning and application of such terms that the apparent simplicity of MCT is stripped away revealing its true complexity" (Airasian et al., 1978, p. 21). In conferences held for the purpose of aiding participants in the identification of competencies, "some participants were surprised and at times disappointed at the lack of consensus regarding answers to such questions as 'what are the definable skills which adults cannot live without?'" (Miller, 1978). Airasian also raises a concern about the particular selection of competencies by suggesting that schools may have promised too much. It is possible, for instance, that schools may have attempted to identify and measure competencies that cannot be achieved by a majority of students, and Airasian asserts that, if this is so, it would be unfair for the schools to expect mastery, and then to penalize students for not achieving it.

The process of setting minimum standards of performance has also been subject to the type of criticism described above in that standards are much more difficult to define and agree upon than might be suspected at first glance. According to a panel established by the National Academy of Education to consider issues on testing and basic skills, "the present measurement arts of educational testing are simply not up to the ambiguous expectations reflected in most state legislation" (NSBA, 1978, p. 13). As with the definition of competencies, an infinite variety of professional disagreement can occur during the identification of minimum performance standards.

Those challenging the second assumption--i.e., that adequate technology exists for measuring competency achievement--point to the problems inherent in validating tests of life skills achievement. A danger already mentioned is that of making competencies trivial in order to render them measurable.

Finally, implementation issues that are raised typically revolve around the methods chosen to solve such problems as what grades to assess, when to apply sanctions for passing or failing the tests, what standards to establish, who should be involved in planning the program and how to promote their involvement, how to deal with students whose native language

is not English, and how to integrate competency testing with the curriculum and with other forms of testing (Greene, 1979). To consider the various answers to these questions and the reasons for particular answers is a major purpose of this document.

Summary

Beyond, or even with respect to, the considerations for and against minimum competency testing discussed above, "school leaders recognize a diverse and contradictory set of motivations: to cut spending and to raise it, to prove schools good and to prove them bad, to cause curriculum change, to help minorities and to legitimize discrimination" (NSBA, 1978, p. 31). There is an old Persian proverb that says: "Where there are two people, there are at least three opinions." That is certainly the case in the controversy over minimum competency testing, and it is also the case that what appears to be an advantage of MCT according to one person is a disadvantage according to another, and vice versa. What can be learned from any discussion then? "The decision of whether or not to implement a minimum competency testing program should involve a weighing of the positive and negative consequences of either decision" (NSBA, 1978, p. 19). Furthermore, it has been urged "that the primary needs people perceive being met by minimal [sic] competency programs be articulated and that these needs be examined in the light of whether such programs, as currently conceived, actually respond to those needs" (Airasian et al., p. 2).

A number of authors suggest, then, that program developers analyze their own needs, consider both sides of the MCT issue in relation to those needs, and also look into possible alternatives to competency testing for meeting those needs. While many advocate using MCT to diagnose students for remediation, for example, it has been suggested that "teachers have probably already identified these students and their problems" (Elford, 1977, p. 10). In addition, "the effective use of test data already collected would seem the most logical approach to early identification and remediation at this level. Local studies could demonstrate the degree to which the elementary achievement tests predict later success in the high school competency test" (Elford, 1977, pp. 10-11). MCT, in other words, may not be the best method by which to collect diagnostic information about students who need remedial aid.

Finally, some authors have suggested that, in making a decision as to whether to implement minimum competency testing, program developers would do well to consider the lead-in time available, the needs of their special student populations, and the funds available. The answers to these questions could determine how feasible minimum competency testing is at a particular time, given that it suits all of the other needs of the developer.

References

Airasian, P., Pedulla, J., & Madaus, G. Policy issues in minimal competency testing and a comparison of implementation models. Boston: Heuristics, Inc., 1978.

American Association of School Administrators. The competency movement: Problems and solutions. Sacramento, California: Education News Service, 1978.

American Friends Service Committee. A citizen's introduction to minimum competency programs for students. Columbia, South Carolina: Southeastern Public Education Program, 1978.

Cohen, D., & Haney, W. Minimums, competency testing, and social policy. Cambridge, Massachusetts: The Huron Institute, 1978.

Elford, G. A review of policy issues related to competency testing for high school graduation. Paper presented at the meeting of the New England Educational Research Organization, Manchester, New Hampshire, May 1977.

Greene, L.F. What can minimum competency accomplish? One response. National Elementary Principal, January, 1979, 23-24.

McClung, M. S. Competency testing: Potential for discrimination. Clearinghouse Review. August 1977, 439-443.

Miller, B. S. (Ed.). Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

National Association of Elementary School Principals. Down and out in the classroom: Surviving minimum competency. National Elementary Principal, 1979, 58(2).

National School Board Association. Minimum competency. A research report, 1978.

Nickse, R. S. Comments on MCT. Proceedings of the National Conference on Minimum Competency Testing. NWREL, October 1978.

Spady, William. Competency-based education: A bandwagon in search of a definition. Educational Researcher, 1977, 6(1).

Scott, L. Summary of the Fall 1978 Conference of the National Consortium on Testing. Cambridge, Massachusetts: The Huron Institute, November 1978.

Walker, D. F. The impact of MCT on school curricula, teaching, and students' learning. Proceedings of the National Conference on Minimum Competency Testing. NWREL, October 1978.

Wilson, H. A. Two sides to test: Positive, negative. National Assessment of Educational Progress, June 1976.

For consideration of a variety of issues related to minimum competency testing and the development of an MCT program, the reader is also referred to:

The Minimum Competency Testing Movement. Phi Delta Kappan, May 1978, 59(9): entire issue.

CHAPTER 2

DEFINING COMPETENCIES

Marcy R. Perkins

Introduction

"Impenetrability! That's what I say!" continues Humpty Dumpty in his discussion on managing words (Lewis Carroll, Through the Looking Glass). In defining competencies, as in defining minimum competency testing, penetrating the wall of words to get through to an acceptable meaning of competencies sometimes seems to be an impossible task. According to the National Council on Measurement in Education (NCME) Task Force, for example, "a review of the elements of competency requirements across state and local districts suggests that the rule for defining competency is that anyone can define it in any way they please as long as they state what they mean" (Bunda & Sanders, 1979, p. 10). The NCME Task Force goes on to assert that no technical definition of competency prevails in the field.

The purpose of this chapter is to consider the issues related to defining competency as a general concept and to competencies as the specific statements forming the basis for measurement in a testing program. Questions about the people who can be involved in competency definition, the processes that can be followed, and the content, format, and organization of competencies that can be specified will all be discussed in this chapter.

One of the principal reasons that so many definitions of competency and so many processes of defining competencies exist is that MCT programs vary greatly in their purpose, size, locus of control, history, and policies. Consequently, it is very likely that one process or set of answers to the relevant issues will not be appropriate for all programs. The intent of this chapter, then, is to bring out the kinds of issues that a program developer or reviewer is likely to encounter, on the basis of situations and occurrences drawn from ongoing programs, and to present potential ways of dealing with the issues, once again on the basis of methods employed in current programs.

Assumptions

While an implicit assumption of this chapter may be that the reader is in some way involved in developing or reviewing a minimum competency testing program, the nature and presentation of the material does not depend upon that assumption. A thoughtful consideration of the issues discussed here might well assist policy makers in making a decision as to whether or not MCT is appropriate for their purposes.

The actual measurement of competencies is the subject of Chapter 3. However, many of the issues that arise during the process of identifying competencies also have implications for how those competencies are measured. Therefore, there will be a certain overlap between this chapter and the next, and it is recommended that both be read for a more complete picture of the activities related to competency assessment.

Limitations

Presenting more than a single process or solution with respect to competency definition does not imply that every one of the possible or existing processes or solutions will be discussed. Limitations on space prevent the fullest treatment of issues possible. Furthermore, this chapter will not discuss competencies in specific subject areas (e.g., reading, mathematics, or democratic governance), nor will it debate the implications of statewide versus local definition of competencies for the meaning of a high school diploma. Rather, processes will be discussed in a general way, as applicable across subject areas and by various governing bodies.

Structure of the Chapter

It is apparent, both in the literature on MCT and in the programs surveyed, that most programs have utilized similar procedures and encountered similar issues in their identification of competencies. In some cases, the process is an explicitly defined one, developed by the agency to facilitate the accomplishment of the task and ensure that the relevant issues are all addressed in some way. In this chapter, discussion will

begin with procedures for competency definition common to programs in the field, followed by the common issues that program developers have faced in implementing those procedures. In conclusion, several examples of overall processes or systems for competency development will be presented and discussed. Before proceeding with the topic of how competencies can be defined, the general concept of competency as it will be treated in this chapter needs to be clarified.

Competency--Competencies: A Treatment of Terms

"Competency" appears to be generally understood in the field as a level of ability at which the examinee can demonstrate the appropriate application of skills to problems or life-role situations (NSPRA, 1978). While the concept of application is not always included in every definition of competency, the notion of a specified or desired performance level is. That is what typically forms the basis of the standards determined for the competency assessment.

"Competencies," by contrast, are seen as specific statements of desired performance. The Northwest Regional Educational Laboratory's conceptualization of competencies, for example, is that "competencies are student outcomes which a school system believes its students should attain before graduation or completion of a course or program" (NWREL, 1978, p. iv). These student outcomes are frequently interpreted as comprising specific learning objectives setting forth those basic academic skills deemed necessary for students to acquire. And these types of outcomes have been called "objectives," "behavioral objectives," "performance objectives," "performance indicators," "standards," "competencies," and other terms, dependent seemingly upon the level of detail and amount of performance specified.

Some take the contrasting view that competencies are different "from other student goals and objectives in that they describe the student's ability to apply basic and other skills in situations that are commonly encountered in everyday life" (NWREL, 1978, p. vi). In Oregon's MCT program, for example, a competency is "a statement of desired student performance representing demonstrable ability to *apply* knowledge, understanding and/or skills assumed to contribute to success in life role functions" (AASA, 1978, p. 45). Still others interpret competencies as being descriptions of common and useful skills, and make no added distinction as to whether these skills are applied, basic, or life-oriented.

While the issue of emphasis (life skill vs. basic skill) related to defining competencies will eventually be discussed in depth in this chapter, the distinction that is necessary at this point is between the generic use of "competency" and the specific use of "competencies." In this chapter, competencies will be used generally to mean specified student learning outcomes. No specific emphasis will be assumed, nor will it be expected that any particular amount of detail is to be defined. These are issues which will be treated within the context of the chapter.

Basic Elements in the Process of Defining Competencies

Defining competencies is a step in the development of a minimum competency testing program that provides structure at two levels in the program. Competencies can be used within the program as the basis for teaching, testing, or both. In fact, all competencies for a K-12 or high school program may be defined, with a subset of these selected for testing at each of the target grades. Defining competencies, therefore, helps to provide the overall instruction/evaluation sequence and scope within the program and to identify the specific domains to be tested. Having a set of competencies is also a prerequisite to determining specifically what the tests will measure, in terms of skills, content, and item difficulty level.

A look at how program developers are identifying competencies in operative minimum competency testing programs indicates the existence of at least eight major steps or components in the process. These include:

-- deciding whether to develop or select competencies;
-- acquiring resources;
-- establishing a task force or advisory committee;
-- developing a competency framework/skill emphasis;
-- defining competency content domains;
-- writing/selecting;
-- reviewing/refining/validating the competencies;
-- selecting the final set of competencies.

Certainly program developers are free to select whatever procedures seem to be most appropriate for their specific programs and particular purposes

and to apply these procedures, with their own staffs or in conjunction with a contractor, in whatever order or process that makes the most sense for their programs. For the purpose of this discussion, however, the tasks will be presented in one possible logical order that has been utilized in the field. Issues related to each of the eight tasks will be raised and discussed, and, wherever possible, alternative activities and ways of applying various procedures will be presented.

Developing versus Selecting the Requisite Competencies

The major difference between selecting competencies and writing them is the source from which they are drawn for inclusion in the program. In other words, those charged with identifying the competencies may nominate ones which they have created or which they have drawn from some extant competency collection. They may also choose to adapt an existing competency rather than to nominate it in its original form. All of the other parameters that must be specified for the competencies, however--such as emphasis, topical domains, number, specificity, etc.--are the same for both selection and development. Similarly, the review, refinement, and validation processes are the same for both.

The decision to develop or select, therefore, may depend entirely on such considerations as the program timelines, resources, and overall goals. Questions to be asked at the outset, then, include:

-- Are the purposes of our program such that we know that there are no competencies extant that will match them?

-- How much time do we have to identify competencies?

-- What is the status of our resources?

If the answer to the first question is affirmative, this will entail the development of competencies specifically geared to the program, which will have the advantage of ensuring a match between program purposes, the competencies, and the assessment of the competencies; it can also engender a sense of ownership in connection with the competencies and the program, because of the high degree of the involvement in this process on the part of the developers. The cost in terms of time and money required to identify competencies, however, are greater for developing than for selecting them; timelines and budgets may therefore preclude the use of this procedure.

If implementation schedules and budgets are restricted, then selecting competencies with their associated assessments may be the more feasible approach. The trade-off here, however, is at the expense of the congruity or fit between the purposes of the program and the competencies selected.

Finally, if it is imperative to identify the competencies immediately and in a constricted timeline, it is most likely that the decision will be to select already existing ones. Since the other procedures involved in identifying competencies apply regardless of whether the decision is made to develop or select, the remainder of this chapter will treat the two together, noting only those points at which they might diverge.

Acquiring Resources

A useful first step in the identification of specific competencies for a program is to review sets of objectives that already exist in the field. Even if the decision has already been made to develop competencies, it is still easier to react to existing materials than to create from a void. And by reviewing objectives in programs similar to theirs, program personnel can begin to define more clearly what kinds of competencies will be needed within the context of their own program. How competencies are identified, and where the skill emphasis should be, can also depend upon the resources available to a particular program. Before such a review can occur, however, resources must be obtained.

The task of acquiring resources is typically undertaken by the program director(s) and can be done even prior to the establishment of a competency testing program. In this case, the existence of appropriate performance outcome statements may well affect the initial decision to implement a testing program.

The ways for obtaining competency statements and sources from which they are available are numerous and varied. For programs concerned with matching competencies and their assessment to existing curricula, lists of skill statements or objectives can be acquired from local schools throughout the state or district. In the North Carolina state MCT program, for example, the final set of reading and mathematics competencies is based upon a collating and ranking of objectives collected from all parts of the state. While this process appears to be straightforward, reviewers are to be forewarned that matching competencies to a diversity of curricula can be no small task.

In other states, lists of competencies that reflect state goals (often with specific performance indicators and sample assessments) are obtainable from the State Department of Education. The Utah Board of Education, for example, appointed subcommittees to develop sample objectives and performance indicators which were to be made available to local districts. Similarly, the California and Illinois Departments of Education provide local districts with technical assistance manuals which include statements of competencies. The advantage of these types of guides is that they present objectives which are matched to state goals and which are perhaps broad enough to be applicable to most curricula within the state. They do run the danger, however, of being too broad in scope to be useful to specific programs.

Finally, commercial objective banks are sources from which competency statements may be drawn. The NWREL has developed a listing of available collections of objectives which is included in the "Outcomes" section of their Guide to Identifying High School Graduation Competencies (1978). In order to help planners select collections that will be most useful to them, this listing provides the following information about each collection that it references: title, description, originator, intended users, purpose/content, usefulness in relation to competency-based education, history of development, related materials, and ordering information.

The acquisition of competencies by one or all of the methods mentioned calls for yet another decision as to who will be responsible for the process of review and for the final selection of competencies for the program.

Establishing a Task Force or Advisory Committee

In general, identifying competencies is accomplished by an advisory committee, often representing a cross section of the state's or district's educators, administrators, and consumers of education (such as parents, students, or business people). In programs that elect to contract with a testing agency for their competency definition and test development, the responsibility of overseeing that work and guiding the development of the competencies still rests with the program staff and/or advisory committee. Frequently, the local or state board of education is responsible for officially adopting or approving a set of competencies, but the first question is how to determine the composition of the set which will be submitted for approval.

While it is not the case that a task force must be established within a competency program, program developers have usually found it advantageous to do so. In general, a greater feeling of ownership in the program, with a subsequent higher probability of program success, occurs when those who will be directly involved in or affected by a program participate in its development.

There are a number of questions to consider, however, in selecting the competency task force. These include:

-- Will there be one group or more?
-- What will the composition of the group(s) be?
-- What will the size of the group(s) be?
-- How will the members be selected?
-- What will be the responsibilities/tasks of the members?

Let us consider each question in turn.

Will there be group one or more? The answer to this question is likely to depend upon other program parameters, such as the overall size of the program and the number of competency areas that have been selected for assessment. If, for example, the competency areas selected are numerous, a task force to concentrate in each area may be desired. Even with as few as two competency areas, separate task forces or subgroups of a larger committee may be sought to represent each of the subject areas.

What will the composition of the group(s) be? The Ohio Department of Education recommends in its Competency Handbook (1978) that competency committees be composed of "administrators, classroom teachers and education specialists." The Colorado Department of Education similarly recommends involving "teachers representing different areas" (Colorado, SDE, 1975), and the Illinois Office of Education (1978) suggests that committee membership might include representatives of the community (e.g., opinion leaders), a cross section of local groups interested in the program, representatives of ethnic and cultural groups, parents, school staff, and students.

It is clear that the options for committee membership are numerous and varied; the ones ultimately chosen may depend on whether the program is to be developed at the state or local level, the resources available for the individual program to draw upon, the kind of expertise desired, and the amount of community involvement desired. The consequences of

passing the competency tests may also help to dictate the composition of the advisory group. If high school graduation depends upon mastery of the competencies, for example, the more defensible the identification process will have to be. Involving a representative sample from different regions of the district or state, from different ethnic backgrounds, from different socioeconomic levels, and from different levels of the educational administration will help to ensure that the process is a legally defensible and politically acceptable one.

What will the size of the group(s) be? The Illinois Office of Education, in Performance Indicators for Competency Assessment (1978), suggests that task forces have 15-25 members, with a number of alternate members, and operating programs have typically had committees of 10-20 members (e.g., Massachusetts, New Jersey, Maryland). Here again, available resources, type of representation desired, and manageability are factors to consider in determining how large a group is desirable.

How will the members be selected? Procedures for the selection of the competency task force committees can include appointment by the state or local board of education; appointment by the local or state superintendent, program coordinator, or other school administrators; random selection by the superintendent or coordinator of representatives of various groups; and open invitation to various groups to obtain their participation. The specific procedure selected, according to the Illinois Office of Education, needs to be "defensible to the public and conducive of efficient task force operations," and "patterned after the selection procedures typically used by the local district for selecting members of other advisory groups" (Illinois, SOE, 1978, p. 9).

What will be the responsibilities/tasks of the members? In general, competency task forces are charged with identifying and recommending a set of competencies upon which to base assessment. Their specific functions will vary depending upon the purposes of individual programs and the reasons for which individual members may have been appointed. For example, members who represent constituencies within the community may provide input from those constituencies to the process of defining competencies. Committee members may also review existing competency sets, review competencies developed by a contractor, and/or develop their own competencies.

Developing a Competency Framework/Skill Emphasis

The basic skills/life skills distinction. Developing a competency framework or skill emphasis really means coming to agreement within the program on the issue of the appropriate context for the competencies. And this relates once again to the primary purposes of the program. Is mastery of the competencies to certify that students possess certain minimum academic skills upon completion of a particular grade (basic skills approach), or is mastery to indicate that students have the skills necessary for adulthood and the situations they are likely to encounter as adults (life skills approach)?

To state it simply, *basic skills* are those skills which parents and society in general expect students to learn and use in school, e.g., reading, writing, and arithmetic. *Life skills* may be these same (basic) skills applied in a "life role" or non-academic context, e.g., reading newspapers (instead of textbooks), filling out job applications (instead of writing book reports), and adding grocery tapes (instead of lists of abstract numbers). Or, life skills may include additional skills not generally considered school skills, e.g., using a telephone, administering emergency first aid, and learning to use a voting machine. A minimum competency testing program, depending on its purposes and emphases, may include any one or all of these approaches to defining those competencies in which students must demonstrate mastery.

The Board of Regents in Rhode Island, as one example, distinguishes among three levels of educational achievement: basic skills, minimum competency, and standards of excellence. Basic skills in Rhode Island comprise specific skills in reading, language arts, mathematics, and cultural arts. Minimum competency, on the other hand, is defined as the achievement of certain basic life skills, or competency in everyday tasks; these tasks are still organized, however, according to the domains of reading, mathematics, language arts, and cultural arts. The standards of excellence, not yet an integral part of Rhode Island's developing program, are considered to be advanced life skills reflected in outstanding scholastic and cultural achievement.

The State Department of Education in Nebraska breaks down the domain of potential skills for assessment in a somewhat different way. In developing the N-ABELS tests, the Department distinguished among life-coping skills, basic skills, and essential learning skills. Life-coping skills are conceptualized in Nebraska in much the same way as minimum competency in Rhode Island; they are defined as those applied skills such as balancing a checkbook and completing a job application. Nebraska's basic skills are similar to Rhode Island's in that they are considered to be skills used

primarily in a school setting. Essential skills in Nebraska, however, are conceptualized as "a subset of the 'basic skills' which are fundamental to continued learning. Essential learning skills are the tools of learning necessary for successful acquisition of competencies in the broader skill areas" (Nebraska, SDE, 1977, p. 1).

In contrast to the programs in both Rhode Island and Nebraska, the Georgia program emphasizes life skill assessment. According to the policy in this state, "the State Board of Education defines as a major role of the public schools the responsibility to ready the children and youth of Georgia for contemporary life roles." The Competency Performance Standards are therefore defined in terms of five life roles: Learner, Individual, Citizen, Consumer, and Producer.

The point which these three programs illustrate is that, while many different labels exist for types of skills, the differences among them are actually superficial. It appears instead that the skill emphasis (life role versus basic) indicates less about the actual skills to be assessed than about the context within which they are to be tested. It may be, in other words, that the same reading skills are invoked when students read textbooks (basic skills) as when they read and respond to newspaper want ads (life skills). Important to keep in mind then, during this process of competency identification, is the relationship between assessment and the way in which competencies are defined.

Basic skills to life skills: issues regarding emphasis. Other issues which the committee may need to consider in specifying an appropriate framework relate to competency measurement, curricula, potential legal problems, and public acceptance.

First, the nature of the definition of a skill or competency will affect the choice of assessment procedures by which to measure student achievement. If the competency is defined as being able to deposit money in a savings account, for example, then the ideal form of assessment is to require a student to go to the bank and deposit a sum of money into a savings account (presumably the student's own). That sort of real-world assessment can be difficult, time-consuming, and costly. A close approximation could be to present students with a simulated deposit situation and require them to fill out bank deposit slips. The result, however, is that test item validity can be called into question, and, indeed, the process of validating the competencies themselves may provide results that can be called into question.

For programs that opt to define their competencies on the basis of their curricula or to structure their curricula to match their competencies, life role competencies can present a problem. Airasian et al.

(1978), for example, question whether we currently have the understanding and technology to teach strictly life-oriented competencies, at least to the degree that we can go "on record assuming the major responsibility for fostering the selected competencies" (p. 29). The NWREL (1978) also points out that the life-role competencies are difficult to identify and agree upon, that they will perhaps require a change in the instructional program, and that they may possibly be so interdisciplinary in nature that curricular change or integration into the curriculum may be extremely difficult to accomplish.

Related to the problems of measurement already discussed is the potential for legal challenge offered by programs in which the competencies are either not directly taught in the curriculum or not established as being valid. Strictly life role competencies are particularly vulnerable to this charge since they can be the most difficult to validate and to incorporate in the curriculum.

Finally, both Airasian et al. (1978) and the NWREL (1978) suggest that the competencies selected for a program should have a broad base of public support, especially given the fact that the impetus for competency testing has come largely from the public sector. Ways of ensuring this kind of support include involving representative audiences in committees and submitting recommended competencies to a general public review.

Following the consideration and discussion of the above issues, one useful approach for the committee to follow from here is to come to a consensus on which general emphasis is desired, and then further delineate domains for assessment. The latter can be accomplished by specifying first those domains to which, in the committee's view, all students have been exposed by the time of testing, and second, those additional domains that represent ideal learning outcomes. Then the committee will be ready to define the content domains for competencies more specifically.

Defining Competency Content Domains

Organization or topical outline. In general, some kind of topical outline by which to organize the competency domains may serve as a useful beginning point for writing or selecting specific competencies. A topic outline or set of goal statements is a general plan for organizing the more specific competencies. In its simplest form, the outline may require only a few category headings to identify subdomains, which can then be

further defined by competencies. For example, a general outline for a language arts test may begin with these categories:

I. Decoding
II. Vocabulary
III. Writing Skills
IV. Reading Comprehension
V. Reference Skills

From this beginning, the outer limits of the domain begin to appear; with each subtopic or competency added under one of the category headings, the shape of the domain becomes more focused.

One consideration to keep in mind here, perhaps, is how the test results are to be reported. When devising a topic outline, programs often find it both convenient and informative to report student scores in terms of domains or subdomains, rather than just by total score for the subject area. Any number of other organizational strategies are also available for generating some kind of topical framework.

It may be the case, for example, that the competency areas, and perhaps some of the specific skill statements, have already been set by the legislature, or by the state or local board of education. In this case, gaps may only need to be identified and filled in, according to the purposes of the program and its relationship to curricula. Other possibilities include taking over a scope and organization from another agency, adopting some form of skill taxonomy, analyzing preexisting curricula and syllabi for an overall framework and determination of scope, adopting a framework identified in national studies, and analyzing the nature and structure of skills typically required in various life roles (NWREL, 1978). In most cases, the purposes and already determined policies of the program can help to determine which approach might be most appropriate.

Since the emphasis in Kanawha County's testing program, for example, is on its interrelation with the curriculum, competencies were identified from the instructional guides and programs already in use in the school district. Both the Arizona and Ohio Departments of Education recommend that competency scope and sequence be linked directly to local district program goals (Arizona, SDE, 1979; Ohio, SDE, 1978). The Illinois Office of Education suggests that local districts identify priority categories for competencies as a first step to selecting or developing them (Illinois, SOE, 1978). And the Colorado Department of Education identifies ways of categorizing objectives in taxonomic domains, with attention to encouraging the development of higher-order cognitive objectives (Colorado, SDE, 1975).

Assessment parameters. After the committee has come to an agreement on a competency framework or topical outline, the members will need to determine the parameters of assessment, since these decisions can help to provide them with guidelines for their actual identification of competencies. Issues for the committee to consider, review, or come to consensus on may include:

-- How many competencies are to be generated?
-- How specific or general are the competencies to be?
-- What are the time limitations on the test?
-- How many competencies per domain should be identified?
-- How many test items per competency will there be?

Frequently, time allocations for testing are predetermined within a program, so that it is the task of program personnel to identify competencies (and later, tests) which can meet their purposes within the specified amount of time. And because of those limits, trade-offs between the number of competencies and the number of items per competency are often necessary. One potential problem to be aware of in making this trade-off is that committee members, because they are concerned about subject coverage, can often be resistant to limiting the skills covered by the test.

There are two possible results of this problem, both presenting some difficulty to assessment and the program as a whole. First, so many overly specific objectives might be defined that dissemination and acceptance of them would be difficult to effect, and assessment options would be restricted. With respect to the latter, for example, suppose under the domain of reading comprehension that the following objective was defined: "The student shall be able to identify the main idea of a newspaper article." This restricts assessment to a particular type of question ("What is the main idea of this article?") and a particular type of item-related content (a newspaper article). Moreover, a case can be made that students should be able to read and comprehend all aspects of movie bills, street signs, advertisements, textbooks, magazines, and various other notices. Either a large number of competencies must be identified to cover all of these circumstances deemed important, or the original competency needs to be made less specific.

The second possible result may be a tendency to make the objectives too general, with the aim of increasing the types of test items that can eventually be matched to them. The problem here is that objectives may be made so general that they will provide no guidelines for appropriate assessment, with the consequence that no reasonable number of items could possibly assess the competency's domain adequately.

With regard to time limits for testing, the total test must be considered when determining both the number of competencies to identify and the number of items to use for testing. According to the state of the art in testing, one minute per conventional multiple-choice item is a general rule of thumb, and a minimum of four items per objective is acceptable in order to meet the minimum for stable reliability estimates (Rubinstein & Nassif, 1977; Schooley et al., 1976). Therefore, within these boundaries, a typical one-hour test can measure 15 competencies. If a longer test is possible, then more competencies may also be specified; if more competencies are necessary or desired, then the effect on test length and time for administration must be considered. While gauging the appropriate level of specificity in order to write or select competencies that can be measured in four or so items is mostly intuitive, practitioners report a surprising degree of agreement when the issues are clearly understood (Rubinstein & Nassif, 1977).

Writing/Selecting

The probable outcomes of the procedures outlined in the previous section are committee agreement on a number of issues (test parameters, competency specificity and scope, competency organization) and perhaps a preliminary specification of a number of competencies. The major task now is to identify those competencies which are probable candidates for inclusion in the final set. As noted previously, committee members can nominate competencies which they have created, or they may nominate competencies from available sources. They may also choose to revise or adapt an existing statement to meet a specific purpose, which is a combination of the two processes.

Regardless of the method used, an initial set of competencies can be generated as a first step. With the parameters and issues in mind that were discussed in the previous section, members may nominate or write, individually or in a group session, any and all statements that, in their view, fit the specifications. Then, review and discussion can occur to settle disagreements about content and to ensure that no gaps remain that need to be filled. At this point, statements that are essentially geared to the same competency may be combined to bring the number of objectives to a more manageable level. Referring to the example described in the previous section, for instance, several competencies may be combined to read: "The student shall be able to read and comprehend materials typically encountered in everyday life (e.g., newspapers, magazines, advertisements, etc.)."

At this point, too, a number of specific issues related to the structure, phraseology, and taxonomic level of the competency statements and the implications of this for assessment come to the fore. First, one standard structure that can be used for generating objectives is Mager's (1962) model, in which an objective has three components: the condition, the performance, and the standard. The condition refers to the given situation to which the performance is related; the performance is the task or skill to be demonstrated; and the standard is the criterion for judging whether or not the examinee has met both the condition and the performance (in the case of multiple-choice items, the standard is always to choose the correct response from the alternatives provided). These components specify the particular parameters within which the assessment can be conducted.

The second issue relates to the nature of the verb that is chosen for any given objective, an issue that is important to consider since the verb will govern the meaning of the objective and the nature of the items that can measure it. Verbs such as "describe" or "discuss," for example, suggest measures other than multiple-choice items; it may be that other types of measurement besides multiple-choice items are desired, but that question is one to consider carefully. Verbs such as "demonstrate" suggest observable performance but do not specify the nature of the performance, and such verbs as "know" or "understand" involve unobservable behavior. Items appropriate for either of these cases are difficult to identify. It is generally advisable, therefore, to select verbs which represent actions that can be tested by the types of items desired for the test.

Finally, the taxonomic level of each competency is a factor to consider when making judgments about the appropriateness of each objective to the grade and skill level for which it is intended. "Taxonomic level" refers to a classification of skills (cognitive, affective, or psychomotor) used to identify the level of, for example, cognitive thought required to demonstrate a particular behavior. Objectives for the third grade, for example, are more likely to deal with knowledge and comprehension, which are relatively simple levels of cognitive process on Bloom's (1956) taxonomy, than with synthesis or evaluation. Verb selection also relates to taxonomic level since verbs can be chosen to reflect specific, desired levels and will influence the type of assessment possible. Verbs such as "define" and "identify," for example, relate to skills at the knowledge level while those such as "apply" and "generalize" can be used in relation to higher-order application skills.

Studying examples of competencies that have been identified for different taxonomic levels, different grade levels, and different purposes is one way in which to gain familiarity with the concepts in order to apply them to the situation at hand.

Reviewing/Refining/Validating the Competencies

Once an initial number of competencies has been generated, either by the selection of existing competencies or the development of new ones or by some combination of both methods, a process of competency review and refinement is generally warranted. In conducting such a review, program administrators may choose to utilize the same staff who identified the objectives, or they may organize a separate review committee (and the same issues of size and selection pertain here as they did earlier). Reviews may be carried out within the committee or agency, or the competencies may be validated through external reviews by the public, other educators, or other professionals. Reviews may be accomplished at meetings or through the use of more formal instruments such as survey questionnaires. How each of these issues is resolved again depends upon the purposes of the individual program, the degree of external or internal approval that is either desired or required, available time, and available resources.

As one example, if input is desired about the relationship of the competencies to skills required in a particular occupational field, then a rating of the objectives by specialists in that field would be appropriate. In this case, questions like the following could be asked about each competency:

-- How often is the skill used on the job?

-- How does the skill relate to emerging fields within that discipline?

-- How important is the skill considered to be, whether it is currently taught or not?

Public acceptance of and involvement in an MCT program may also be obtained through external reviews of the competencies by citizens, teachers, parents, and representatives of the business community. In this case, the questions may be of a broader nature, particularly if the competencies are intended to reflect life-oriented skills (e.g., "Do you think ninth-graders should know this?").

Whether or not extensive public surveys are selected as a means to competency validation, reviews by content specialists are frequently conducted to ensure content validity of the competencies. Additional committees of content specialists may be formed for this purpose, or locally or nationally known specialists may be asked to react to the materials.

Table 1 presents examples of the types of criteria that can be adapted and utilized in programs for such reviews. These criteria represent the types of judgments that programs typically make about skills statements, whether the judging is accomplished formally or informally.

Selecting the Final Set of Competencies

On the basis of the results of reviews conducted, the competencies can be refined and finalized. At this point, another round of reviews may be conducted, or additional input may be solicited from various groups if the need for either is felt. If not, then the final product is complete and ready for implementation with the competency testing program.

Summary Guidelines for Defining Competencies: Three Examples

Whether a choice is made to follow an explicit model such as those presented in this section or to define a unique process and set of procedures will depend upon the particular program--its purposes, timeline, resources, staff experience, and staff interest. The purpose of this section is to provide additional resources upon which administrators may choose to draw.

Presented briefly are frameworks for competency definition established by the Illinois Office of Education, the Ohio State Department of Education, and the Northwest Regional Educational Laboratory.

Illinois

Presented in Figure 1 is a process for defining competencies that was constructed by the Illinois Office of Education as a resource for local districts which opt to implement minimum competency testing. In Illinois,

TABLE 1

Criteria for Reviewing Competencies

Teachability

(1) Is it possible for the schools to teach the knowledge, skills, and/or attitudes described in the competencies?

(2) Is curriculum available related to the individual competencies?

(3) Will remedial programs, if needed, be available now or in the near future?

Acceptability

(1) Do the competencies represent reasonable standards of proficiency to be required of all students?

(2) Are the competencies agreed to be necessary outcomes for student success in school or their daily lives?

(3) Are the competencies reasonable, appropriate and important outcomes of the total educational experience?

Bias

(1) Are the competencies free of statements that suggest that some social, occupational or life roles should be valued more than other roles?

(2) Are the competencies free of bias related to sex, race, age, region, religion, ethnic, or cultural background?

TABLE 1 (continued)

Generalizability

(1) Are the competencies achievable regardless of students' sex, socioeconomic status, race, rural or urban setting, and religious belief?

(2) Will all students for whom a particular competency is applicable be exposed to sufficient instruction to achieve the specified knowledge, skill or attitude?

(3) Are the competencies appropriate for those students who transfer within the state?

Suitability

(1) Are there available and acceptable ways to measure the outcomes specified by the competencies?

(2) Can the competencies be measured within the schools' time constraints and resources?

(3) Will adequate educational resources (e.g., time, staff, money, books and materials) be made available to support the implementation of the competencies now or in the near future?

(4) Are the competencies free of specifications which would require special equipment or facilities which are not available to most students?

Validity

(1) Does the content of the competency fit the intent of the topic or goal statement for which it was written?

(2) Can items be developed for the competency to measure the domain intended by the topic?

TABLE 1 (continued)

(3) If competencies are to be used as a basis for promotion or graduation, do the competencies represent levels of student proficiency and accomplishment of sufficient importance?

(4) Does each competency identify a significant or important skill, in relation to the infinite number of skills which could be chosen?

Specificity

(1) Are the competencies worded specifically enough so that it is clear what skills are and are not included in the competencies?

(2) Is the content domain specified by the competency too broad or too narrow?

(3) Is each competency unique, mutually exclusive, so that extensive overlap does not occur among them?

Taxonomic Level

(1) Is the taxonomic level of the competency appropriate to the subject matter of the topic and to the grade level?

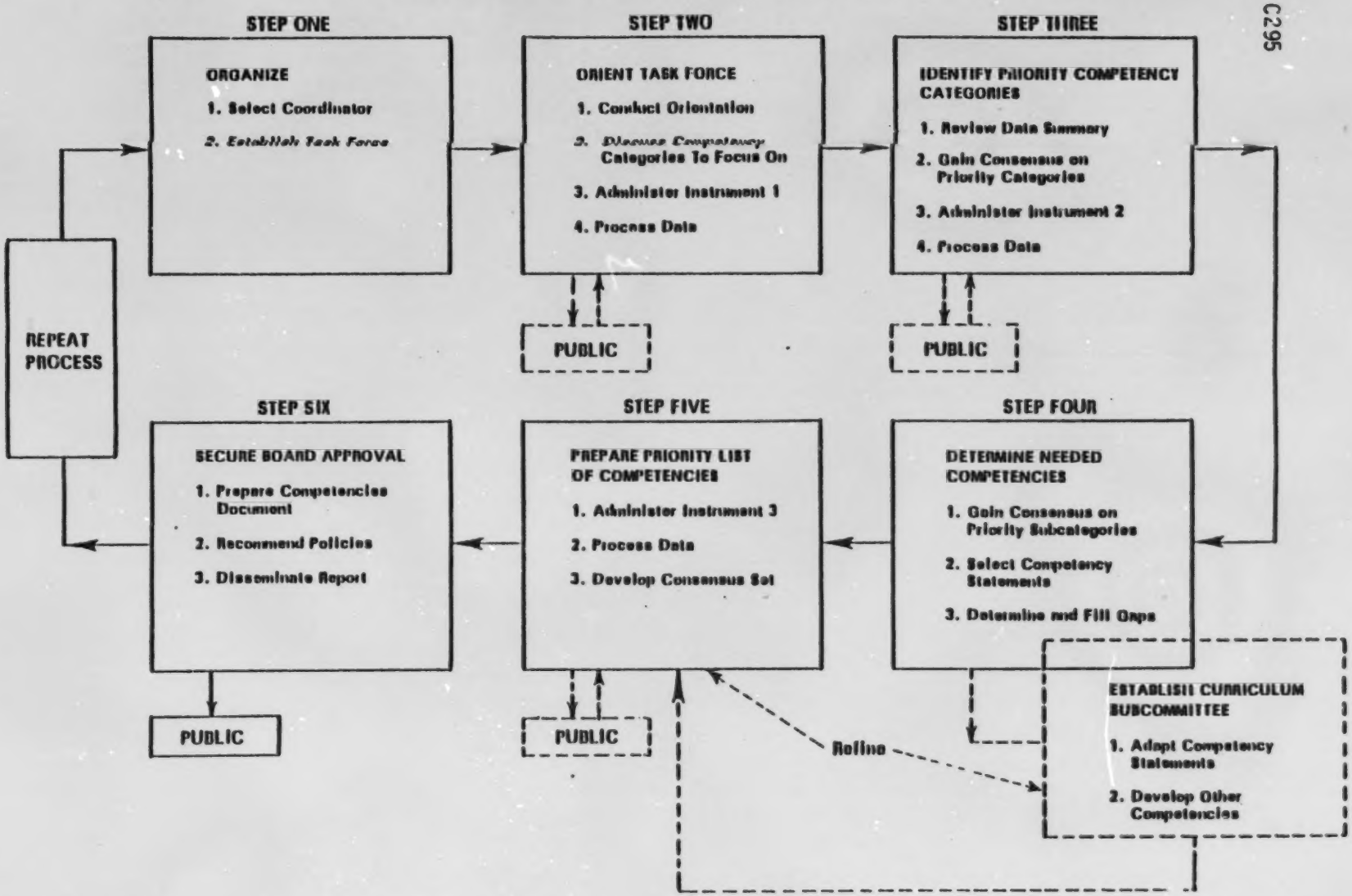
Clarity

(1) Are the competencies free from jargon, slang, colloquialisms, or other unusual terms?

(2) Are the competencies stated clearly and succinctly?

(3) Are the competencies written so they communicate effectively to students, parents, community members, teachers, administrators, and other interested individuals?

FIGURE 1*
Defining a District Set of School Leaving Competencies — The Process



* From Performance Indicators for Competency Assessment, Illinois, State Office of Education, Springfield, Illinois: Author, 1978.

MCT is a local district option, and the state offers technical assistance in the form of published documents, consultants, and regional centers to the local districts selecting the option.

As evident in the figure, Illinois has identified six basic steps leading to the identification of a set of competencies suitable for graduation assessment. The first involves the selection of a competency coordinator and the members of a task force who will be directly involved in carrying out the process. During the orientation of the task force, competency areas are ranked in order of importance to the individual school district. Then in Step 3, competency category priorities are established and specific competency subcategories are rated. When priority subcategories are established, initial competency statements are selected and others are developed to fill whatever gaps are noted. Each statement is then rated by asking how important it is that a student acquire this skill before leaving high school. On the basis of the results, a priority list of competencies is established. The last major step in this process is to present the competencies to the district board of education and obtain approval for implementation. It should be noted, too, that the system allows for public involvement at a number of points as well as for a cyclical process of refinement.

Ohio

The Ohio Department of Education and the NWREL both take a more question- or issue-oriented approach to establishing a competency program and defining the requisite competencies. The Ohio Department of Education recommends that a task force or advisory committee consider the following questions carefully in their competency identification process:

-- What is the purpose of the competency program?
-- What competency areas will be addressed?
-- What grade levels should be used for measurement?
-- Shall there be individualized or uniform competencies?

The purpose of considering these issues is, in the view of the Department, to help provide a framework for administrative decision making.

C295

NWREL

The series of questions that the NWREL considers in its Guide to Identifying Graduation Competencies (1978) are:

(1) What is a graduation competency?

(2) What kinds of knowledge, skills and attitudes should be included in graduation competencies?

(3) How can one determine that the coverage of a set of graduation competencies is accurate?

(4) How general or specific should the content of a competency be?

(5) What degree of difficulty should graduation competencies represent?

(6) Should the same set of graduation competencies be adopted for all students?

(7) Who should be involved in drafting and adopting a set of graduation competencies?

(8) What format should be used for stating graduation competencies?

With this set of questions as a basic framework, the NWREL discusses the nature of competencies in a competency-based educational program, their potential role in the educational system, and what their adoption as a graduation requirement can mean in terms of measurement, instruction, and instructional management.

References

Airasian, P., Pedulla, J., & Madaus, G. Policy issues in minimal competency testing and a comparison of implementation models. Boston: Heuristics, Inc., 1978.

Arizona, State Department of Education. Suggested guidelines for the development and implementation of a continuous uniform evaluation system. Phoenix, Arizona: Author, 1979.

American Association of School Administrators. The competency movement: Problems and solutions. Arlington, Virginia: Author, 1978.

Bloom, B. S. Taxonomy of educational objectives: Handbook I, cognitive domain. New York: Longmans, 1956.

Bossone, R. M. Minimum competencies: A national survey. New York: City University of New York, Center for Advanced Study in Education, 1978.

Bunda, M. A., & Sanders, J. (Eds.). Practices and problems in competency-based measurement. NCME, 1979.

California, State Department of Education. Technical assistance guide. Sacramento, California: Author, 1977.

Candor-Chandler, C. Competency measurement at the local level: A case study of Kanawha County Schools, West Virginia. In R. B. Ingle, M. R. Carroll, & W. J. Gephart (Eds.), The assessment of student competence in the public schools. Bloomington, Indiana: Phi Delta Kappa, 1978.

Colorado, State Department of Education. A school improvement accountability process. PAK #3.1: Writing student objectives. Denver, Colorado: Author, 1975.

Illinois, State Office of Education. Performance indicators for competency assessment. Springfield, Illinois: Author, 1978.

Mager, R. F. Preparing instructional objectives. California: Fearon Publishers, 1962.

National School Public Relations Association. The competency challenge. Arlington, Virginia: Author, 1978.

Nebraska, State Department of Education. Nebraska--Assessment battery of essential learning skills administrative manual (rev. ed.). Lincoln, Nebraska: Author, 1977.

Northwest Regional Educational Laboratory. A guide to identifying high school graduation competencies. Portland, Oregon: Author, 1978.

Ohio, State Department of Education. Competency handbook. Columbus, Ohio: Author, 1978.

Schooley, D. E., Schultz, D. W., Donovan, D. L., & Lehman, I. J. Quality control for evaluation systems based on objective-referenced tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

CHAPTER 3

TEST SELECTION AND DEVELOPMENT

Michael Priestley

Introduction

Purpose

The entire process of selecting or developing a test is analogous to distilling salt water: you begin with a barrel of brine, turn it into steam, filter out the impurities, condense it, and out the other end comes a quart of water pure enough to drink. In selecting or developing a test, a similar process takes place: the process, from beginning to end, is one of filtering, refining, and gradually defining the material in specific terms so as to produce a test that is good enough to use.

The purpose of this chapter is to discuss issues and procedures related to test selection and development that may be useful to practitioners who are responsible for planning, implementing, or reviewing the test selection and development components of a minimum competency testing program. The site visits conducted in this study of MCT programs, as well as an analysis of materials disseminated by the programs visited, revealed a set of basic issues of concern to practitioners in the field, and it is this set of issues that forms the basis for this chapter. In addition, the program personnel who were interviewed identified various procedures for selecting and/or developing a test that they had found useful.

Although specific issues and procedures are presented here, these are neither exhaustive nor prescriptive. Rather, they represent lists of concerns and practices cited by program personnel. While, as in distilling water, a certain sequence of events is a prerequisite for manufacturing pure water, there is still some latitude with respect to both the order of events, and within each task, the procedures used to complete it. Similarly, although the activities discussed below are presented in a certain order, a sequence which is based upon that followed by many programs, it is not the case that the tasks associated with test selection and development must be sequenced in this order. Both the order of the tasks and the means used to carry them out are choices to be made on an individual basis, taking into account the specific circumstances of each program.

Content

The first section of this chapter covers the issues related to determining whether to (1) select a test; (2) develop a test; or (3) combine the two approaches in some fashion, for example, by selecting half of the test items and writing the other half specifically to match the identified competencies. These issues represent concerns voiced by personnel in the field. The following sections of the chapter describe procedures for each of the options that either have been or are being used by existing programs. The final section discusses issues relating to test validity and reliability as identified by practitioners and measurement experts. When appropriate, examples involving existing programs will be cited and relevant materials disseminated by programs referenced.

Context

In its basic philosophy, this chapter is intended to present a nonjudgmental view of the process of selecting and/or developing a measurement instrument to be used in any type of minimum competency testing program. It is important to stress at the outset that this chapter does not advocate any one method or approach over another, but only presents information which will assist educators in making informed decisions. As much as possible, however, the advantages and disadvantages of each option will be noted in an effort to provide a complete and sound basis for a decision. Similarly, when testing programs are cited in the course of discussion, the only purpose of these citations is to illustrate an issue or procedure, or to substantiate a general statement, and not to praise or criticize any particular program.

The considerations in selecting and developing tests are so numerous that a treatment of this size and scope cannot possibly cover all of them. This chapter will, however, attempt to present in detail the issues considered most important by those who have faced them in the field, and to provide lists of resources which may be of use to anyone who wishes to investigate further aspects of test selection and development which may only be mentioned briefly here.

With regard to the specific context of this chapter, two basic points must be made. First, the reader will note that the chapter deals only with the selection and/or development of criterion-referenced tests. This

focus follows from the finding that all programs surveyed use criterion-referenced instruments. In one program, a norm-referenced test was altered for use as a criterion-referenced test, a procedure that will be described later in the chapter.

Second, an important point must be made in relation to the three standard domains into which competencies are most often divided: the cognitive, affective, and psychomotor domains. This chapter will focus primarily on the cognitive domain, which includes the skills normally required for such subjects as reading, writing, mathematics, science, and history. The affective domain, which includes competencies related to attitude, personality, and emotional behavior, will not be treated in depth. There has been some debate on the issue of whether a minimum competency program should properly test in the domain of the affective competencies, and testing in this domain was far from widespread in the programs surveyed for the study. For opposing views, see Wilson (1978) and Ahmann (1978).

In this chapter, as stated, the emphasis is on the cognitive domain; the affective and psychomotor skills will be discussed briefly in touching upon other competencies which might be tested in addition to those competencies of which students must actually demonstrate mastery. Testing in the affective and psychomotor domains, however, may be useful for diagnosis and remediation, and/or in relation to data analysis of test scores for the cognitive domain.

There are certain specific issues and procedures which this chapter cannot explore in depth, because of its size and scope and because of its stated intention (to provide a basis for making decisions in any minimum competency testing program). One of these issues is the development of tests for special populations such as the physically or emotionally handicapped, or people whose first language is not English. Since it is not possible here to state general guidelines which would be applicable and useful in these situations, it is perhaps best to leave this complex matter in the hands of local administrators, who are the persons most familiar with the needs and requirements of their own special populations.

Initial Decision: To Select or Develop

Preliminary Considerations

Both the experience of practitioners and materials published on this topic suggest that program planners consider various questions prior to deciding whether to select or develop a test. Brickell (1978) reviews some basic questions concerning competency measurement, while the California Department of Education in its technical assistance manual (1978) addresses the decisions of whether to select or develop a test. Resolving the following points may facilitate both the decision-making process and the implementation of the decision:

- IDENTIFYING THE PURPOSE OF THE TEST. A minimum competency test may be designed for any number of purposes, including diagnosis, screening, evaluation, certification, and application/ advancement/selection. With the purpose of the test clearly defined, subsequent decisions related to selecting and developing a test can promote the effort to match the test with its intended purpose.

- IDENTIFYING WHO IS TO BE TESTED AND WHEN. This step can help a testing program administrator to plan the program in response to the stated purpose of the test. For example, if a minimum competency test is to certify students' mastery of competencies as a requirement for graduation, should students be tested at the end of the twelfth grade? the eleventh grade? the beginning of the ninth grade? These decisions will significantly affect the test itself, the mode of administration, and many other issues.

- IDENTIFYING THE DOMAIN OF THE TEST. In the previous chapter it was pointed out that the choice of domain can have far-reaching effects on the nature of the competency test. A *domain*, as used here, is the universe of content knowledge and skills which are defined by the competencies, and which the test will measure. *Competencies*, as used in this guide, are statements of behaviors or skills which the examinee must demonstrate by his or her test performance, e.g., "Identify the definition of a vocabulary word in the context of a sentence." (Such skill statements are

variously called objectives--performance, behavioral, assessment, or terminal objectives; performance indicators, or performance expectations.) The domain identified reflects the purpose of the test. For example, if the purpose is to certify mastery of basic skills (reading, writing, and mathematics) among third-graders, then the domain is defined by competencies that third-graders can reasonably be expected to have achieved in these subject areas.

Which competencies are identified and how they are stated will determine, in part, what kind of test will be required. If the competencies state that a student must demonstrate the ability to write a coherent paragraph, then that much of the test has been determined: all students tested will be required to write actual paragraphs. Similarly, if the purpose of a test is to certify auto mechanics, then one of the competencies may state that a candidate must be able to change a flat tire. This will then determine that part of the test must require candidates to perform actual functions required of auto mechanics.

Issues to Consider

There are many issues to consider in deciding whether to select or develop a test. The Ohio Department of Education (1978), for example, raises the issues of timeline and the availability of commercial instruments that measure the competencies of interest. Five issues in particular were of concern to program planners who were responsible for deciding whether to select or develop their competency tests. Each one is discussed below, on the assumption that the primary goal is to develop some sort of preliminary test specification, a blueprint or descriptive plan of the test, from which the initial decision can be made.

Consequences to examinees. Major decisions made on the basis of results from a minimum compe- tency test may determine whether or not a student needs remediation, whether or not the student should be promoted or graduated, and whether or not a candidate will be certified or licensed as a professional (e.g., a firefighter, a teacher, a veterinarian). The more serious the consequences to the examinee, then, the more important is the reliability of the information upon which the decision is based. If students are to be denied diplomas, for instance, on the basis of a test,

that test should certainly be a valid and reliable instrument. (A later section in this chapter treats the technical issues of validity and reliability.)

Moreover, the more serious the consequences to examinees, the more likely is the possibility of a legal challenge. A candidate who has been denied a license to practice architecture because of failure on a certification test may have a legal right to challenge the validity and the reliability of the test.

The quality of the instrument is one basis for legal challenge. Merle McClung (1977), of the Education Commission of the States, formerly of the Center for Law and Education in Cambridge, Massachusetts, identifies several bases for legal challenge, which are summarized below.

- VALIDITY. A minimum competency test used to make decisions regarding the remediation, promotion, or evaluation of students must have "curricular or instructional validity," i.e., it must test what students have actually been taught.

- BIAS. A test used to make decisions regarding students, job candidates, etc., must not have an adverse impact on any minority group (EEOC, 1977); this includes the perpetuation of "prior effect" of racial discrimination, e.g., tracking. Even a test that is proven to have curricular validity may cause adverse impact; if the curriculum is biased and the test measures the curriculum, then the test may have adverse impact.

- PHASE-IN PERIOD. McClung states that a test designed to measure 12 years of cumulative knowledge which is implemented in a phase-in period of two years is unfair to students who then have only two years to prepare for the test. And the decision in a recent case in Florida (Debra P. v. Turlington, 1978) conforms with this view.

Domain and competencies. The second issue is a pragmatic one: the domains and competencies identified for testing will influence the feasibility of selecting a prepared test. If the competencies require a paper-and-pencil test to measure reading and mathematics skills, then there are many tests available which may be appropriate. The further the domain diverges from the basic skills, the more difficult it may be to find a test which measures that domain and its identified competencies. Few

tests, for example, measure a student's ability to paint a sign, or use a voting machine, or analyze the logic in a political debate--all of which could be considered minimum competencies.

Most practitioners will agree that the ideal test provides a direct measure of competencies, but not all programs can accommodate the costs, the time required, and the other problems connected with obtaining the best of all possible tests. Possible compromises include using indirect measures of competencies, e.g., a multiple-choice test of writing skills instead of an actual writing sample. A competency that requires a student to write a grammatically correct sentence may be measured by having the student pick from four sentences the one that is correct, or--more indirectly--the one sentence out of four that is incorrect.

Again, validity is an important requirement: the test selected or developed must measure the specified competencies, which are most often based on the curriculum. The question of direct versus indirect measures of those competencies may be relegated to secondary importance if the selected test measures the competencies in some way. Validity between a selected test and the competencies varies with the subject area: tests of basic skills such as reading and mathematics are most often based on standard curriculum and the most basic competencies. For other subjects such as health and physical education, nutrition, speech, social studies, and economics, it is likely to be more difficult to find a published test which measures the competencies identified in one of these areas because of the lack of agreement as to what skills constitute the basic competencies in these subjects. A developed test, on the other hand, can be constructed specifically to measure the competencies identified in almost any area.

Timeline. Developing a test usually requires more time than selecting one. The National Assessment of Educational Progress (NAEP), for example, spends about two years in the development of one of its tests (AASA, 1978). The time devoted to selection or development will vary depending upon a number of factors. Administrators and teachers in South Burlington, Vermont prepared tests to measure all of the state competencies during the summer of 1977 and revised the instruments the following year. This was made possible by the state-mandated schedule of implementation.

In some cases, however, legislative mandate has required the implementation of testing programs within only a few months. In New Jersey, for example, the mandate called for immediate implementation of minimum competency testing. Although Minimum Basic Skills tests were to be developed, given the schedule of implementation, the tests developed for the

statewide assessment program were used during the first year. These instruments were replaced the following year by the newly developed Minimum Basic Skills tests.

A shorter time is usually required to select a test than to develop one because the process is much simpler: a selected test requires no item writing, for instance, and often requires only limited field-testing, if any, before actual administration.

Resources. The cost of developing a test is generally greater than the cost of selecting a test, since test development involves not only out-of-pocket expenses but also the cost of staff time. The amount of staff time allocated varies according to who actually does the developmental work, such as writing items. Staff may be employed only as project monitors to coordinate the volunteer efforts of teachers, or they may be responsible for developing the entire test in-house. In the latter case, if permanent full-time staff are used on a particular project, additional personnel may be needed to assume the responsibilities of the regular staff for the duration of the project.

The cost of administering and scoring tests may not vary significantly between the selected test and the developed test, but administration costs vary with the program design. If a test can be administered by local teachers, for example, this is less expensive than establishing a special team of people to administer all tests. The latter procedure is used in many programs to improve test security and to standardize conditions for test administration, as in New Jersey where the administration of basic skills tests is supervised by county test coordinators.

In programs in which part or all of the project work is awarded to a consultant, the cost of the consultant must be considered. Costs associated with consultant contracts are generally higher for test development than for test selection because of the likelihood that more expertise will be required in developing a test than for advising on the selection and use of a published test.

Availability of technical expertise. Expertise is usually required both for selecting and developing tests, although to different degrees. Types of expertise required most often involve psychometric issues such as test reliability, validity, scoring, and data analysis; but specialized knowledge may also be needed in such areas as curriculum, subject matter, and readability.

To select a test, technical expertise will be required in assessing the quality of test instruments, i.e., their validity, reliability, and lack of bias. This expertise will be required to develop criteria for screening tests and to develop procedures for applying these criteria systematically to the potential tests.

To develop a test, technical expertise will be required in order to write or select, review, and edit test items, as well as to sequence the items and to actually construct the test.

Whether the decision is to select or develop a test, expert knowledge will be advantageous in planning and implementing the overall design of the testing program and of the test, since someone familiar with these processes will be aware of the consequences and ramifications of each step in the process. For example, a knowledge of how the test will be scored, e.g., by specific subtests or by competencies, is essential in determining the number of items that must be developed for the test. Whether the scoring provided by the test publisher (if it is provided) will suit the needs of the program, i.e., tell you what you need to know, may also influence the decision to develop or select.

All of these issues are ones to consider in deciding whether to select or develop a test. Once these issues have been resolved, a picture of what the ultimate test may look like will begin to form.

Preliminary Test Specifications

At this stage, program planners may want to consider preparing *test specifications*, a kind of blueprint which serves as an ideal description of the final test. Test specifications are useful at this point because they can bring out issues and concerns that may not have been considered. For example, if a test is selected or developed this year and administered to ninth-graders, should the same test be used next year? Or must parallel or equivalent forms of the test be developed now for later use? Obviously, the answers to these questions will affect every aspect of the test design, down to the choice of distractors in each test item.

In the programs surveyed, the test specifications were generally developed as part of a total program design to ensure that the test which was selected or developed would meet the needs and purpose of the testing program. Careful consideration can then be given to test length, the number of competencies measured, and the numbers of items required to

yield the types of scores desired (e.g., by subtest or by competency). These decisions must be made in relation to all of the factors discussed previously, but particularly in relation to the purpose of the program (what kinds of test scores are needed?); cost (how much more will a 100-item test cost than a 50-item test? how much more will it cost to score open-ended items or writing samples than multiple-choice questions?); and time, in relation to both the time required to implement the program and the actual time for administering the test (how much of the student's and teacher's time should be spent on this test?).

This rough form of the test specifications, then, usually includes: (1) an estimate of how long the test should be; (2) how many competencies will be measured and by how many items; (3) what kinds of items will be used; and (4) what kinds of scores the test will generate. After this the decision of whether to select or develop the test may become clearer, as the outline of the test emerges from these specifications. This preliminary test design may have to be modified later, particularly if the decision is in favor of test development; but this step offers a glimpse of the chimera being stalked in the search for the right test.

Types of Tests

The final issue to consider independently before making the decision is what type of test to use, i.e., given specific competencies, how will these be measured? Program planners can turn to various programs and materials for examples of different kinds of assessments. The California Department of Education discusses various modes of assessment, including performance-based testing, in its resource materials prepared for local districts (1978).

Three general approaches to testing are available: performance tests, observational tests, and paper-and-pencil tests. A performance test can be defined as one "that measures performance on tasks requiring the application of learning in an actual or simulated setting" (CAPT Newsletter, 1978). This type of test is the most direct method of assessment in a number of situations: when testing competencies that require physical manipulation, such as using a telephone in an emergency situation (e.g., ConVal, NH), building with blocks, performing calisthenics, or adjusting a carburetor; or when testing competencies that require "on-the-job" situations involving social interaction, such as sales techniques, giving and following oral directions, or making introductions (e.g., South Burlington, VT).

Observational tests measure competencies which involve behaviors or skills that can only be assessed through observation, e.g., teaching skills, or behaviors in a social setting. This approach to testing is not entirely divorced from performance testing (obviously, if an examinee is performing, someone else must be observing in order to evaluate the performance), but observational testing also has unique characteristics in that it can be conducted in real or simulated situations. Observational tests are especially useful in situations which are not contrived intentionally for the purpose of testing, e.g., observing preschool children for such competencies as attentiveness, observing rules of conduct, and initiating conversation with peers.

Paper-and-pencil tests are by the far the most common tests in use; these include multiple-choice tests, tests made up of open-ended items (such as fill-in-the-blank items), and essay-type tests (which include writing samples, design problems, and so on). Brickell (1978) contends that paper-and-pencil tests can be considered performance tests in a school situation because taking a paper-and-pencil test is an actual performance in school. These tests essentially measure the application of knowledge and skills.

Different kinds of tests generate different kinds of results, so the choice of what type of test to select or develop is a very important one. To some extent, the type of test required is determined by the competencies to be measured. For example, the most direct assessment of a competency that is stated as "Describe the four basic food groups" would be an oral or written description. Each type of test is ideally suited to particular competencies. One type of test may be chosen to measure all competencies, with some measured more directly than others. Or some combination of the types of tests may be chosen: a performance test for life skills, such as comparison shopping or using the library, and a paper-and-pencil test for the school skills. The state of Hawaii, for example, has developed a battery of tests for third-graders to measure 100 different competencies (termed performance expectations). The test battery, used for screening and then for diagnosis, includes hundreds of test items of different types: performance items for physical exercises, oral-response items, verification checklists and rating scales, and several types of paper-and-pencil items. Hawaii's educators have chosen to use the most direct method of assessment in every case possible, on a competency-by-competency basis.

In general, cost and objectivity favor the use of paper-and-pencil tests, but relevance and face validity may favor the performance test (Brickell, 1978; Mehrens, 1978). The further one goes from performance tests and the closer to paper-and-pencil tests, in general, the less expensive testing becomes. Also, the closer to paper-and-pencil tests, the larger is the number of the tests one will have to choose from.

Making the Decision

This is the point at which all of the factors, issues, and considerations discussed can be weighed, one against the other. The relative importance of cost versus direct assessment, of timeline versus legislative mandate, of test validity versus available resources, and so on, can be determined with respect to the specific program under development. The results of all the preliminary analyses of these issues and considerations can be helpful in making the decision.

The following two sections of this chapter describe actual procedures generally followed in selecting or developing a test, or to achieve a combination of the two. The first of the two sections deals with test selection, the second with test development. Following these sections, the discussion returns to issues which affect both developed and selected tests, such as field-testing, technical and legal issues of validity and reliability, and test administration.

Test Selection

Programs choosing to select a test (e.g., North Carolina) typically carry out a number of procedures prior to that selection. These include considering the test domain and sources of possible tests, developing criteria for selection, identifying potential instruments, and applying the selection criteria in order to arrive at a decision. In considering selection criteria, program planners may elect to use criteria that have already been developed and used, such as the MEAN System developed by the Center for the Study of Evaluation (CSE). Developing or choosing criteria for selection as well as the other procedures will be discussed later in this section.

Considering Test Domain and Sources of Tests

Domain. The key element here is congruence, i.e., the relationship between what the test purports to measure and the competencies that have been identified for testing. Madaus et al. (1979) state that congruence is a function of two considerations: (1) the number of competencies measured by the test, and (2) the number of items measuring each competency. An initial review to identify tests which measure both the broad competency areas (e.g., reading, mathematics) and the specific competencies, with an appropriate number of items measuring each competency, may narrow down the number of tests to consider as potential candidates.

Program planners may encounter difficulty in finding a test which measures exactly those competencies identified for testing in a particular program; a reasonable approach, therefore, is to seek the test(s) measuring the largest percentage of those competencies.

With regard to the number of items required per competency, Berk (1979) states that the number varies in relation to four essential factors: (1) importance and type of decisions to be made on the basis of results; (2) relative importance assigned to the competencies; (3) the number of competencies; and (4) practical constraints. Berk recommends that 5-10 items per competency be used for most classroom decisions and 10-20 items be used for school, system, and state-level decisions. More items per competency will be required for scoring by competency--i.e., determining pass/fail or mastery for each skill--than for scoring by subtest or total test score. Fewer items may suffice in certain situations, as in a test for which there is to be only one total score, or one score for each of two or three subareas; then the total number of items on the test or in each subarea outweighs the importance of the number of test items per competency. The number of items must be considered carefully in relation to the criteria listed above to ensure selection of a valid and reliable test.

Sources of tests. Sources from which instruments may be selected include normative-referenced and criterion-referenced tests, with corresponding competencies specified, and item pools or banks, large sets of items from which appropriate measures may be selected. (For a list of test sources, see Appendix A.)

As soon as potential "candidates" for use in the testing program have been identified, planners may write to the publishers of the tests they wish to acquire and request copies of the test itself, answer keys, technical manuals, and any other information that may be helpful. In addition, planners may wish to follow the example of the Massachusetts Committee on

Basic Skills Improvement Policy (Madaus et al., 1979), and ask the test publishers to identify those items which, in their opinion, measure the competencies already identified for the program.

Developing Criteria for Selection

This task usually involves working with staff and other persons (e.g., teachers, parents, students, business representatives, members of the community, legislators) to develop a comprehensive list of criteria by which to judge the available tests. The particular selection criteria and the method of review that is chosen or developed will vary according to program needs.

A number of methods for reviewing tests for local, state, and national programs are discussed in the program and research literature on this subject (see Appendix A). In Massachusetts, for example, the Department of Education contracted with the Public Affairs Research Institute to develop both criteria for screening commercial tests and a system for applying these criteria. (Madaus et al., 1979). The Institute identified criteria relating both to the content of the test and its technical properties. The MEAN System developed by the Center for the Study of Evaluation (CSE) is another example of possible selection criteria. The acronym stands for the four characteristics a test is rated on: measurement validity, examinee appropriateness, administrative usability, and named technical excellence. For a discussion of how the CSE staff applied this system, see CSE (1976).

Criteria to consider include not only technical (e.g., validity, bias) and content-related (e.g., accuracy, difficulty) issues, but also practical features such as cost, availability of tests, and the administration of test instruments. To facilitate this step, a comprehensive set of review criteria can be developed to match the needs of the specific program.

Identifying Potential Instruments

This simply consists of weeding out those tests which are obviously irrelevant or have obvious flaws. For example, tests of personal attitudes or civic responsibilities are irrelevant if the identified competencies

cover only reading and mathematics. Also, a test of mathematics which requires only logical thinking and no computation may be considered inadequate if computation is a specified competency.

Applying Selection Criteria

Once the potential instruments have been collected, the people appointed to review the tests can do so by applying whatever selection criteria have been developed or chosen. There are a number of ways in which this step may be accomplished. Most of these procedures involve the use of rating scales or checklists to quantify data from the review process. The planner may wish to consult the test evaluations conducted by CSE using the MEAN system. The evaluations of commerical tests for all levels are listed in Appendix A.

It is important to note that in situations which involve lay people as reviewers--e.g., parents or legislators who may not be at all familiar with testing programs--program personnel have found it generally advisable to train these people before the actual review process begins, in order to guarantee the internal consistency across reviewers which is essential to the review process.

If there is more than one committee, then different selection criteria may be developed for each committee. Or different people may rate the tests on the basis of one portion of a complete set of criteria: for example, in Massachusetts (Madaus et al., 1979) the committee was composed of technical experts who reviewed the tests for technical criteria, and other members, primarily teachers, who reviewed the tests on the basis of content criteria. Although practical issues were of secondary importance in Massachusetts because the tests selected were not mandated, but only approved for use, a different set of people might be appointed to review the tests only in terms of practical concerns. Ultimately, the results from all of these separate reviews will be aggregated.

Given a committee of people appointed to review the potential tests (preferably the same people who developed the selection criteria, who will thus be familiar with all aspects of the program), there are at least three approaches to completing the review. One approach is to have everyone rate each test independently and compile the results through a totally objective method, e.g., keypunching and computer analysis. A second approach, which may be favored by people who wish to feel personally involved in a group process, is to have all the reviewers evaluate each

test simultaneously as a group, keeping one record of the consensus on each test. A useful compromise between these approaches is to have the reviewers rate the tests independently, and then meet as a group to discuss their ratings and reach a committee consensus.

The advance preparation of rating scales or checklists which are easy to read, understand, and use is well worth the time and trouble required. The Competency Handbook (Ohio SDE, 1978) provides a number of models which include a checklist of purposes for measuring competencies, a list of criteria for test selection, a test nomination form, a test selection information form, a test comparison grid, and a rating scale for determining the relative importance of all the selection criteria.

A sample section of a rating scale from the Ohio Competency Handbook (1978) is included below.

INSTRUMENT SELECTION CRITERIA

Directions: Each committee member should sort the criteria into one of three categories. "H" is highest priority or most important. "M" is medium priority and "L" is lowest priority. The entire committee should then adopt a consensus list of criteria. IT IS MOST IMPORTANT TO CONSIDER EACH OF THESE ITEMS IN TERMS OF THE PURPOSES SELECTED EARLIER.

Circle One:

H M L 1. Cost per student including materials and desired scoring services.

H M L 2. Total amount of time necessary for test administration.

H M L 3. Ease of administration (e.g., can be given by teachers.

H M L 4. Recent appropriate norms (i.e., for different times of year and for groups of students similar to yours).

H M L 5. High reliability and validity for the purposes in local testing program.

Careful planning and preparation can keep the difficulties in reaching consensus to a minimum. The more explicit the selection criteria and the more practical and efficient the review procedures, the easier it will probably be to reach agreement as to which tests are most appropriate for use in the program.

After each reviewer has rated each potential test on the basis of the set criteria, and/or the group has reached a consensus on each test, the results can be compiled and analyzed.

Selecting the Instrument

For this final step, it may be helpful to rank-order the tests which have received the highest ratings. This will be easy to do if the rating process yields a summative score for each test: for example, the MEAN system, which yields a four-letter rating of each test ("Good" ratings on each of the four criteria used in this system would be recorded as "GGGG"). Other procedures may yield numerical scores if a certain number of points are awarded for the adequacy of the test in relation to specific criteria.

When the top tests have been rank-ordered, there may be some with equal or very similar ratings. In this case, the group can return to the individual ratings and weigh the pros and cons of the results of the review in order to eventually reach consensus on which test is most appropriate for the testing program. If after rank-ordering the potential tests there is one test rated high enough that it stands head and shoulders above the crowd, then the job is done: a test has been selected.

Test Development

In the programs surveyed test development typically proceeded in one of three ways: (1) by constructing it from the ground up, which includes writing all of the items; (2) by selecting the items from item pools or from other tests, or by modifying an existing test; or (3) by using a combination of the two methods by writing some items and selecting others. Regardless of which approach is taken, procedures that can help to ensure that the test meets its intended purpose are similar. These are:

-- identifying personnel to develop tests;
-- developing test specifications;
-- developing item specifications;
-- writing/selecting items;
-- reviewing and editing items;
-- field-testing the instrument;
-- conducting validity review;
-- modifying the test, if necessary.

Each will be considered in turn.

Identifying Personnel to Develop Tests

Depending on the scope of the test development project, a number of personnel may be needed to complete the process. These include item writers, editors and reviewers, test administrators, project coordinators, content experts, and technical experts to assist in designing, scoring, analyzing, and reporting the results of the test. Whether personnel with these skills are to come from local school districts, department staff, or consulting agencies will depend on the individual situation; advantages and disadvantages exist for each possibility.

Teachers may be called upon to perform many of the developmental tasks, particularly test item development, validity review, and test administration. In relation to the task of item development, programs have found both advantages and disadvantages to using local teachers. According to Miller (1979), many teachers have had little or no training in evaluation and find it difficult to develop good test items. On the other hand, teachers have a vested interest in developing a test that may affect their schools: they may make up for lack of expertise with their enthusiasm and willingness to learn. Training teachers to develop test items can be beneficial to students, to school systems, and to the teachers themselves; it may also save money if teachers are willing to contribute their time and energy in return for training and experience. A test developed by local personnel is more likely to receive strong support in the schools than a test developed elsewhere.

In Peterborough, New Hampshire, administrators hired two district teachers and sent them to Educational Testing Service (ETS) to learn how to write and edit test items. Although prior to this experience they had no specialized training in this area, the teachers assisted in the development of the competencies and assessment, and now administer and score competency assessments at all grade levels. By comparison, South Burlington administrators chose to develop all instruments in-house and provided training in a summer workshop for interested teachers. Those participating received credit toward advancement on the district's salary schedule.

Another option for administrators interested in providing staff members with special knowledge in test development is to bring in consultants who will train district staff. In Gary, Indiana, for example, consultants from ETS taught teachers how to score essay tests holistically.

Test development experts are often more helpful in those areas which require technical expertise, such as designing the test and analyzing the results--expertise which is less likely to be available in the school

districts or in a department. The budget, however, may determine whether consultants are employed in a test development project. Some compromises may be possible; districts may opt to develop some tests on their own and contract for others. Although Gary, Indiana developed an oral proficiency test and hired consultants to teach the scoring of an essay test, the reading test was developed by Westinghouse Learning Corporation, which selected items to match the competencies.

Finally, the program staff generally assumes the responsibility for monitoring and coordinating personnel, and completing the activities involved in the project.

Developing Test Specifications

As mentioned in an earlier section, a blueprint or set of specifications for the test is helpful before construction of the test begins. A test developer who builds a test without a preliminary design faces the same risk as an architect who begins construction of a house before drawing the blueprint: the structure may collapse. For examples of test specifications, see California's Technical Assistance Guide (1978), Appendix B. Sample test specifications from materials prepared by the Beaumont Unified School District are included in this document.

When designing the test, an important consideration is the domain covered by the test. Each subarea of the domain is usually to be represented on the test in proportion to its importance in the domain. For example, a possible domain may be mathematics, which is to be measured in a one-hour test. Then within the domain there may be subareas such as mathematical computation, number concepts, geometry and measurement, and problem solving. When the relative importance of each subarea within the domain has been determined (by public survey, job analysis, committee consensus, etc.), each subarea is represented proportionally on the test.

On the next page is a sample chart that may be useful in constructing test specifications; the numbers and competencies have been devised to describe a mathematics test of 60 items in length. Note that in this hypothetical domain there are four subareas, each of which is tested by a number of items proportionate to its predetermined relative importance. Within a subarea, also, the importance of each competency has been determined, and assigned a number of items needed to measure it.

C295

DOMAIN: MATHEMATICS				
Subarea	% of Domain	Number of Items/ Subarea	Competency	Number of Items/ Competency
I. Computation	20%	12	1.	4
			2.	8
II. Concepts	30%	18	3.	6
			4.	6
			5.	6
III. Geometry & Measurement	10%	6	6.	2
			7.	4
IV. Problem Solving	40%	24	8.	6
			9.	8
			10.	10
TOTALS	100%	60		60

One purpose of drawing up test specifications is to obviate problems that might arise in the future. The types of scores expected and the decisions to be made on the basis of test results will help to define the test specifications. Within the subareas of a particular test, the number of items matched to each competency must also be determined. In general, the larger the number of items, the greater the reliability of the test results and the greater confidence one can place in evaluation decisions (Berk, 1979). Berk recommends 5-10 items per competency for classroom-level decisions and 10-20 items for school-, system-, and state-level decisions, and this was generally observed in the field.

When the test specifications have been completed, each subsequent step in the development process can readily follow from these specifications.

Developing Item Specifications

Test specifications are to tests as item specifications are to items: they help you plan in advance to determine what the items will look like. As Dahl (1971) and Rovinelli and Hambleton (1976) point out, the most important requirement in the construction of a criterion-referenced test is establishing a direct relationship between each item on the test and the competency it purports to measure. Translating competencies into items is an essential step in establishing the validity of the test, and a carefully designed framework for this process can significantly improve chances of success (Priestley & Nassif, 1979). Item specifications can be of tremendous value in this process because they can determine exactly what types of items must be written or selected to measure each competency. The process of reviewing these specifications may also, as in New Jersey, for one, serve to promote confidence that an appropriate and useful instrument is being constructed.

Generally, specifications for selected-response items include most or all of the following characteristics: a statement of the competency, a sample item, stem attributes (how the question is to be presented), and response attributes (how the distractors are to be constructed). They may also include stimulus attributes (description of stimulus material the item requires, e.g., the length and difficulty level of a reading passage) and a description of the content domain (e.g., what subjects can be tested across a set of items matched to the competency).

For constructed-response items such as essay tests or oral-response questions, the specifications usually include a description of the testing situation (e.g., "The student will listen to questions on a paced-tape and record his or her responses"), and a set of scoring criteria to determine whether or not a response is adequate.

The ideal situation is to have the same people who first developed the item specifications then write or select the items. Through discussion, close examination of the competencies, and attempts to define them more clearly, the specification developers can acquire a detailed knowledge of just what each competency entails and how it should be measured. Once the type of measurement and its characteristics have been defined, item development may proceed.

The development of item specifications, as stated above, is an optional step; many test development procedures go directly from competencies to items without it. Whether or not this extra step is taken will depend upon factors considered in this section, particularly those related to test validity. If the validity of a test is paramount, then item specifications will help to ensure that the process of item development will generate a valid test. Item specifications may also save considerable time and money by averting a situation in which the first administration of a completed test reveals that the test does not measure what it was intended to.

Writing/Selecting Items

To construct a test, items may all be written, may all be selected, or may be generated through a combined approach.

If the test items are selected, two considerations arise: (1) availability of sources from which items can be selected, and (2) how these items will be matched to the competencies already identified. Items can be selected from item pools or from published tests--usually tests in the public domain. Complete item pools, or item banks, do not yet exist, although they are under development in many areas (AASA, 1978). In some cases, local school districts and consortia have pooled their resources to develop item banks, and some states (e.g., California, New York) are in the process of developing pools of items for use by their local districts. Commercially developed item pools are also available. (See Appendix A for sources for items and tests.)

Selecting items from published tests, which Florida chose to do to obtain some of its items, may incur additional costs for permission to use the items (some publishers charge a standard rental fee per item, per administration), or it may simply require a request for permission to reprint copyrighted materials and an acknowledgement to the publisher cited in the test booklet. The fact that tests in the public domain do not require permission for reprinting may be a trade-off for lower quality; items are released after they have been used, and may sometimes be outdated.

The second important factor in selecting items is establishing congruence between what the items measure and what the competencies intend to measure, i.e., matching items to competencies. The most widely used approach to this task is a review by a committee of qualified persons, often teachers and evaluators. The first step is to define the exact intent of each competency; next, to identify the content or skill measured by each item reviewed; last, to match the items to the competencies. This review can be performed on a group or individual basis, in much the same way as the review for selecting published tests.

Although this approach sounds simple, it presents some difficulties in that criteria other than item/competency match must be applied. For example, item bias is a concern; ensuring that the set of items matched to one competency covers a representative sample of skills or knowledge in the domain defined by the competency can be difficult; and equating difficulty levels of items measuring a single competency can be a problem. (For a more comprehensive list of criteria for item review, see "Reviewing Items.")

If the decision is to write the items required to construct a test, an important consideration again is item/competency match. The use of item specifications is one effective way to help ensure that valid items will be produced. The likelihood of producing valid items will be greater if the item writers are qualified content specialists with demonstrated "minimum competency" in writing skills, and if they have been carefully and systematically trained by a professional experienced in item writing, editing, and item writer training. Those with little or no experience in item writing may need a practical introduction to evaluation concepts before they begin writing. Also, if the team of writers is a representative sample with respect to ethnic background, sex, and cultural perspective--insofar as this is feasible--this may decrease the possibility of bias in items across the test. (For sources of information on how to write test items, see Appendix A.)

Other procedures which may be followed to ensure the validity and quality of test items take place after the items have been written, at least in first draft form. These procedures are described in the next four subsections.

Reviewing and Editing Test Items

Although the entire set of items is likely to be reviewed many times in the course of test development, and it is useful to have as many people as possible participate in the review process, for the very first review it may be more desirable to submit the items to a two-member team consisting of a qualified reviewer and a qualified editor who can examine each item for content and grammar, respectively, as well as for the quality of the item as a measurement device. The team is likely to be most effective if both reviewer and editor have been trained in certain aspects of psychometrics and test development.

Items are generally reviewed on the basis of many criteria. Some of these criteria, compiled from a number of programs, are listed in Table 1, and may be used for reviewing all types of items, either written or selected. For a listing of some additional guidelines for reviewing and editing, see California (1978), Appendix C. For the purposes of clarity, item-related terms used below are defined as follows:

DIRECTIONS:	Instructions used to orient examinees to the item format. How to answer the question(s).
STIMULUS:	A reading passage, picture, chart, etc., that includes information necessary to the item.
STEM:	The main body of the item which states any necessary facts and asks the actual question.
ALTERNATIVES:	Possible answers to choose from, which often include a correct response and one to four distractors.
DISTRACTOR:	An incorrect response in a set of alternatives or possible answers.

Once the test items have been reviewed and edited, they may still need revision by the item writer, or they may have to be replaced by new items. When a set of acceptable items has been produced, they can then be reviewed by a committee of qualified persons who did not participate in the writing process. The most widely used approach is to select a committee that comprises teachers, one or more members of the departmental

C295

TABLE 1

Item Review Criteria

Validity

- Is the item closely matched to the competency, i.e., does the item measure knowledge or a skill within the domain of the competency?

- Is the knowledge or skill measured by the item a significant aspect of the domain (which may contain an almost infinite number of possibilities)?

- Is the format of the item suited to the skill or knowledge it is intended to measure?

Bias

- Could the item be more difficult to one group than to another because of an unstated assumption or esoteric wording? Is the item biased in terms of sex, race, age, culture, religion, or region? Does it contain a stereo- type?

  (NOTE: Bias is a multi-level criterion, i.e., each item can be reviewed individually and as part of the entire set of items. One item which presents a woman as a secretary or a Chinese man working in a laundry is not necessarily biased; if similarly stereotyped situations occur in many items, then the test as a whole may be biased.)

- Could the item be offensive to a member of any ethnic group?

TABLE 1 (continued)

Accuracy

- Is the item grammatically correct?
- Is there only one correct response?
- Are the stem and alternatives clearly stated and unambiguous?
- Are there structural clues to the correct response?
- Are all distractors plausible but still incorrect?

Difficulty

- Is the readability level of the item appropriate to the grade level?
- Is the level of difficulty of the skill or knowledge required by the item appropriate for the designated grade level?
- Are all items matched to each competency geared at about the same level of difficulty?

TABLE 1 (continued)

Interest Level

- Will the item-related material, e.g., reading passage, be interesting to examinees?

- Is there enough variability in approach and content across items, to the extent that variability is possible, to make the test interesting?

Practical Considerations

- Is the item simply too big and unwieldy to be included, e.g., an item which requires a student to choose one of four maps when a question about one map would suffice?

- Is the format of the item clear and understandable?

- Are the directions clear, concise, and unambiguous?

staff, and perhaps parents, student representatives, members of the business community, legislators, etc. The selection of this group may depend on the importance of the test and its consequences, the amount of time and money available, and the nature of the test itself (i.e., the subject matter, as in a reading test versus a test in engineering). As large and representative a cross-section of people as possible, given project constraints, is desirable.

It may be necessary to orient committee members to the testing program and train them in how to interpret the competencies and review test items. Results from a committee consensus on each item written will often determine the amount of revision necessary before field testing.

Field-Testing the Instrument

Program personnel identified during the site visits two procedures as being particularly useful in developing a competency-based instrument: (1) field-testing the instrument with a representative sample of the population for whom the test is intended, and (2) examining the test and its results to determine the effectiveness of the test in measuring what it purports to measure. The first procedure--field-testing--is considered here. The second involves determining validity and reliability, which is discussed in the next section.

The purposes of a field test usually include one or more of the following: (1) to refine the test items; (2) to identify "bad" items, i.e., items which do not yield the kinds of results expected; (3) to obtain baseline data for assessment; (4) to obtain data for designing the final test(s), e.g., data from which to construct parallel or equivalent test forms; and (5) to gather information useful for refining the instrument as a whole, e.g., time required for administration. Field-testing is that step in developing a test which will generate the results from which evaluative decisions may be made. Field-testing may also be required for selected tests, for one or more of the purposes stated above.

When the field test results have been collected and analyzed, their interpretation can be used to refine the test items, to provide a basis for final test design, and to provide empirical data for selecting items to be used in the final test.

Conducting Content Validity Review

In recent years content validity review has become increasingly important, particularly in relation to tests which are used for certification, licensing, and high school graduation. Whether or not this step is taken, however, will depend on the situation.

Essentially, a group of content specialists in a particular field review the test items to determine whether or not they are content-valid. On the basis of the ratings of specialists, an item is classified as valid or not valid. (See the next section, "Establishing Validity and Reliability," and Chapter 4, "Standard Setting," for further discussion of rating procedures.) Results of a content validity review can be used to refine the items; e.g., an item declared invalid in the first review may contain only a minor error that can be emended.

If a test does not require the time and expense of this somewhat complicated, technical process, e.g., if the test is intended for use as a classroom instrument for fourth-graders, then a content review by specialists in a particular field may be considered. This may be helpful in several ways: a specialist may discover something in the test which is outdated as the result of a recent discovery (e.g., the planet Pluto is temporarily not the farthest from the sun; thus a science test item becomes invalid). Or the expert may notice something that was overlooked in several reviews by staff members who have been closely connected with the project. A disinterested eye can often spot flaws that go unnoticed by others. Also, review and a "stamp of approval" from recognized specialists can give the test a great deal more credibility in the eyes of the public and of other professionals.

Modifying the Test, If Necessary

The final step in developing a test is to modify the items and the test design itself, if necessary, on the basis of the results of the item reviews, the field test, and the content validity review. When this step has been completed, the test can be prepared for printing, distribution, and/or administration.

Establishing Validity and Reliability

Of the many technical and legal issues related to minimum competency testing, the validity of the instrument used to certify attainment of the competencies is generally agreed to be one of the most important. A close second is the issue of test reliability. The steps required for the process of either test selection or test development may be determined by the need for establishing validity and reliability. The importance of these issues bears a direct relationship to the seriousness of consequences to the examinees and the likelihood of legal challenge. A test required for high school graduation, teacher certification, or professional licensure is much more likely to be challenged on legal grounds than, for instance, a test used to diagnose reading difficulties among third-graders. In Florida, the SSAT-II, for example, which is a requirement for high school diplomas, has been challenged in the courts (Debra P. v. Turlington, 1978); as a result, it was necessary to establish validity of the instrument and its reliability as a basis for decisions related to the attainment of competencies. For a test that is susceptible to legal challenge, technical assistance from test developers, measurement experts, and legal experts may be desirable.

Types of Validity

The purpose of this section is to provide some practical definitions of technical terms, a discussion of ways in which these issues might affect a minimum competency testing program, and suggestions for procedures which can be used to establish the validity and reliability of a test instrument.

When a challenge to a test arises, the courts generally rely on the widely accepted Standards for Educational and Psychological Tests (APA, AERA, NCME, 1974) as the authoritative source on such issues as validity and reliability. The Standards recognize three major types of validity, which are defined below.

Content validity requires that the skills, knowledge, and behaviors measured by a test constitute a representative sample of the skills, knowledge, and behaviors in the performance domain. Critical components of content validity include the clear definition of a performance domain, of the competencies on which the test is based, and of the method for sampling from the domain.

Construct validity refers to the ability of a test to measure the constructs or intellectual concepts which it is designed to measure. Examples of such constructs are reading readiness, management aptitude, and attitude. Establishing construct validity requires one or more predictions about the hypothetical characteristics of examinees who score high on the test as opposed to those who score low, and data with which to prove the validity of these predictions.

Criterion-related validity includes both concurrent validity and predictive validity. *Concurrent validity* consists of establishing the validity of an instrument by analyzing it in relation to a concurrent criterion, e.g., a student's grades or scores on an existing test already proven valid. *Predictive validity* requires the demonstration of a correlation between performance on the test and degree of success in relation to the predictor, e.g., college entrance or job performance.

In addition to these types of validity defined by the Standards, three other types are often mentioned. These include the following: *curricylum validity*, which demonstrates the degree to which a test measures what is purportedly taught in the schools (often considered part of content validity); *instructional validity*, which demonstrates that students have actually been taught what is on the test (often considered part of criterion-related validity); and *face validity*, a nontechnical, informal term that implies that a test looks valid, i.e., appears to be a reasonable measure of the desired competencies.

Construct validity, according to Linn (1979), is useful in minimum competency testing if the inferences made from the test results lead to expectations about the examinee's aptitude, e.g., a student who passes a test in addition is now ready to begin learning subtraction. This type of validity, however, is difficult to achieve and seldom practical for application to achievement tests developed by educators to determine mastery/nonmastery (Nassif, 1978).

Federal guidelines state that a test for licensure or certification can be considered valid if it can be shown that the test measures a representative sample of the skills required in the performance of the job for which a candidate will be licensed or certified (EEOC, 1977). Both criterion-related validity and content validity are considered acceptable means by which to establish job-relatedness. This concept of job-relatedness is analogous to the requirement that a minimum competency test used in an academic setting must measure the skills and knowledge required of a student to perform in school or after graduating from the school.

Content validity can be used to support the inference that a person who passes a test based on a clearly defined domain and made up of a representative sample of items measuring that domain has attained at least

some degree of competency in relation to the skills and knowledge identified. This is the most practical approach to validating a test based on school skills. A test based on life skills or survival skills, however, may require the use of predictive validity to show that results from a test a student has taken in school have a definite relationship to success in life beyond the schoolyard.

McClung (1977) contends that the most important types of validity for a minimum competency test will more often be curriculum and instructional validity. A test may be challenged on a sound legal basis if it does not measure what students are taught in school; a test may be challenged, therefore, if it measures the stated curriculum but that curriculum is not actually taught in the classroom.

Types of Reliability

Reliability refers to the degree to which the results of testing are attributable to systematic sources of test score variance (Standards, 1974). In other words, a test is considered reliable if it generates comparable test scores across time, across test forms, and/or across subareas of the test's domain.

Reliability is particularly important in relation to the generalizability and consistency of inferences made on the basis of test scores, e.g., mastery/non-mastery (Hambleton & Novick, 1973; Linn, 1979). This is perhaps the characteristic of reliability that is most relevant to minimum competency testing.

Characteristics of reliability include one or more of the following elements: comparability of test forms, which refers to the consistency of scores across different forms of the test designed to be parallel, or equivalent; internal consistency, which refers to the correlation of scores between test halves or subtests in a test battery; and comparability over time, which refers to the reproducibility of scores on a test given more than once.

In a minimum competency testing program, several situations may arise which would increase the desirability of ensuring test reliability. For example, if a test is given to high school students, some of the students may fail, receive additional instruction or remediation, and then take the same test again or take a second form of the test. The student's scores on each test must be consistent with respect to the student's achievement

in order to produce reliable results. Also, a minimum competency test may comprise subdomains, such as reading and mathematics, and/or subareas within the domains, such as problem solving and computation. The level of difficulty across subareas and the reliability of scores achieved in these subareas must be established.

Procedures for Establishing Validity and Reliability

Certain accepted procedures for establishing the reliability and validity of a test seem applicable to minimum competency testing; they will be described here briefly. Further assistance in these procedures can be obtained from the extensive literature extant on these subjects and through the use of professional testing specialists.

Establishing Validity

The most important type of validity in a minimum competency program is likely to be content validity, i.e., establishing that the test measures the specified domain of competencies. This is also the most practical type of validity procedure to conduct in terms of cost, time, and usefulness of the results, particularly in a public school testing program. As mentioned earlier, content validity procedures stand up in court as acceptable if done correctly.

Probably the most widely used approach to determine content validity is the review of the test items by a group of at least 10 content specialists. Each content specialist reviews each item independently on the basis of four criteria: item/competency match (whether the item measures the competency, and whether the entire set of test items constitutes a representative sample of all the competencies in terms of their relative importance); significance (whether each item measures a significant aspect of the domain of a competency); bias; and accuracy (whether there is only one correct response). In addition to these criteria, a content validity review may incorporate an analysis of the level of difficulty of each item and its appropriateness on a minimum competency test (see Chapter 4, "Standard Setting").

Each reviewer should be trained in the rating procedure, and then instructed to rate each item individually as valid or not valid, on the basis of the established criteria. Results from all of the independent reviews are collected and analyzed for consistency across reviewers, and then presented in summary form for interpretation. A consensus of the reviewers is necessary to establish the content validity of an item; however, the number required for consensus will vary with the situation (see Nassif, 1979). All items rated as valid are then available for use on the final test.

Establishing Reliability

Reliability is determined on the basis of empirical data collected from actual test administration(s). This can be done by administering the same test to the same examinees at two different times (with too little time in between administrations to allow significant learning by the examinee); administering two parallel forms of the test to the same examinees; or administering two halves of the test at once, providing that both test halves are representative of the same domain.

Different methods of estimating reliability are designed to account for different sources of measurement error (Standards, 1974). As a general rule, the longer the test is (in terms of the number of items), the more likely it is to be reliable. In many minimum competency testing situations, however, the test is not usually of sufficient length to ensure reliability on this basis alone (Linn, 1979).

Appendix A

Sources

Tests: published

Buros, O. K. (Ed.). The nineteen thirty-eight mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1972. (Originally published, 1938)

Buros, O. K. (Ed.). The nineteen forty mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1972. (Originally published, 1941)

Buros, O. K. (Ed.). The third mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1949.

Buros, O. K. (Ed.). The fourth mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1953.

Buros, O. K. (Ed.). The fifth mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1959.

Buros, O. K. (Ed.). Tests in print. Highland Park, New Jersey: Gryphon Press, 1961.

Buros, O. K. (Ed.). The sixth mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1965.

Buros, O. K. (Ed.). The seventh mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1972.

Buros, O. K. (Ed.). Tests in print II. Highland Park, New Jersey: Gryphon Press, 1974.

Buros, O. K. (Ed.). English tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Foreign language tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Intelligence tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Mathematics tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Reading tests and reviews II. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Science tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Social studies tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Buros, O. K. (Ed.). Vocational tests and reviews. Highland Park, New Jersey: Gryphon Press, 1975.

Tests: unpublished

Johnson, O. G. Tests and measurements in child development: Handbook II. San Francisco, California: Jossey-Bass.

Johnson, O. G., & Bommarito, J. W. Tests and measurements in child development. San Francisco, California: Jossey-Bass, 1971.

Test items

Instructional Objectives Exchange
P.O. Box 24095
Los Angeles, California 90024

National Assessment of Educational Progress
700 Lincoln Tower
1860 Lincoln Avenue
Denver, Colorado 80203

C295

Sequencing

Test items

Ahmann, J. S., & Glock, M. D. Evaluating pupil growth: Principles of tests and measurements (5th ed.). Boston: Allyn and Bacon, 1975.

Bloom, B., Hastings, J. T., & Madaus, G. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Gronlund, N. E. Measurement and evaluation in teaching (2nd ed.). New York: Macmillan Co., 1971.

Henrysson, S. Gathering, analyzing, and using data on test items. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Review

Tests

California, State Department of Education. Technical assistance guide for proficiency assessment. Sacramento, California: Author, 1977.

Center for the Study of Evaluation. CSE Elementary Test Evaluations. Los Angeles: University of California, 1970.

Center for the Study of Evaluation. CSE ECRC Preschool/Kindergarten Test Evaluations. Los Angeles: University of California, 1971.

Center for the Study of Evaluation. CSE RBS Test Evaluations; Tests of Higher-Order Cognitive, Affective, and Interpersonal Skills. Los Angles: University of California, 1972.

Center for the Study of Evaluation. CSE Secondary School Test Evaluations. Los Angeles: University of California, 1974. (three volumes)

Center for the Study of Evaluation. CSE Elementary School Test Evaluations. Los Angeles: University of California, 1976.

Madaus, G., Airasian, P., Hambleton, R., Consalvo, R., & Orlandi, L. Development and application of criteria for screening commercial standardized tests for the Massachusetts Basic Skills Improvement Policy. Boston: Public Affairs Research Institute, 1979.

National Consortium on Testing. Testing the tests (Staff Circular No. 1). Cambridge, Massachusetts: Huron Institute, 1978.

Ohio, State Department of Education. Competency handbook. Columbus, Ohio: Author, 1978.

Writing

Test items

California, State Department of Education. Technical assistance guide for proficiency assessment. Sacramento, California: Author, 1977.

Gronlund, N. E. Measurement and evaluation in testing (2nd ed.). New York: Macmillan Co., 1971.

Ohio, State Department of Education. Competency handbook. Columbus, Ohio: Author, 1978.

Wesman, A. Writing the test item. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

References

Ahmann, J. S. Basic issues concerning competency-based testing. In R. B. Ingle, M. R. Carroll, & W. J. Gephart (Eds.), The assessment of student competence in the public schools. Bloomington, Indiana: Phi Delta Kappa, 1978.

Airasian, P., Pedulla, J., & Madaus, G. Policy issues in minimal competency testing and a comparison of implementation models. Boston: Heuristics, 1978.

American Association of School Administrators. The competency movement: Problems and solutions. Arlington, Virginia: Author, 1978.

Berk, R. A. Some guidelines for determining the length of objective-based criterion-referenced tests. Paper presented at the meeting of the National Council on Measurement in Education, 1978.

Brickell, H. M. Seven key notes on minimal competency testing. In B. S. Miller (Ed.), Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

California, State Department of Education. Technical assistance guide for proficiency assessment. Sacramento, California: Author, 1977.

Candor-Chandler, C. Competency measurement at the local level: A case study of the Kanawha County Schools, West Virginia. In R. B. Ingle, M. R. Carroll, & W. J. Gephart (Eds.), The assessment of student competence in the public schools. Bloomington, Indiana: Phi Delta Kappa, 1978.

Dahl, T. Toward an evaluative methodology for criterion-referenced measures: Objective-item congruence. Paper presented at the meeting of the California Educational Research Association, San Diego, 1971.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hathaway, W. E. Competency measurement at the local level: A case study of the Portland, Oregon Public Schools. In R. B. Ingle, M. R. Carroll, & W. J. Gephart (Eds.), The assessment of student competence in the public schools. Bloomington, Indiana: Phi Delta Kappa, 1978.

Linn, R. L. Issues of validity in measurement for competency-based programs. In M. A. Bunda & J. Sanders (Eds.), Policies and problems in competency-based measurement. NCME, 1979.

Madaus, G., Airasian, P., Hambleton, R., Consalvo, R., & Orlandi, L. Development and application of criteria for screening commercial standardized tests for the Massachusetts Basic Skills Improvement Policy. Boston: Public Affairs Research Institute, 1979.

McClung, M. Competency testing: Potential for discrimination. Clearinghouse Review, August 1977, 439-443.

Mehrens, W. A. The technology of competency measurement. In R. B. Ingle, M. R. Carroll, & W. J. Gephart (Eds.), The assessment of student competence in the public schools. Bloomington, Indiana: Phi Delta Kappa, 1978.

Miller, B. S. (Ed.). Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

Nassif, P. M. Standard-setting for criterion-referenced teacher licensing tests. Paper presented at the meeting of the National Council on Measurement in Education, Toronto, March 1978.

National Council on Testing. Testing the tests (Staff Circular No. 1). Cambridge, Massachusetts: Huron Institute, 1978.

Ohio, State Department of Education. Competency handbook. Columbus, Ohio: Author, 1978.

Pasch, M. Minimal competency testing: The problem of validity. Paper presented at the meeting of the American Educational Research Association, San Francisco, 1979.

Priestley, M., & Nassif, P. M. From here to validity: Developing a conceptual framework for test item generation in criterion-referenced measurement. Educational Technology, 1979, 19(2), 27-32.

Rovinelli, R. J., & Hambleton, R. K. On the use of content specialists in the assessment of criterion-referenced test item validity. Paper presented at the meeting of the American Educational Research Association, San Francisco, 1976.

Standards for educational and psychological tests. Washington, D.C., 1974.

U.S., Equal Employment Opportunity Commission. Guidelines on employee selection procedures. Federal Register, 1978, 43(166), 38290-38309.

CHAPTER 4

SETTING STANDARDS

Paula M. Nassif

Introduction

The purpose of standard setting is to specify the score above which performance is considered satisfactory and below which it is considered unsatisfactory and thus to characterize the capabilities or competencies of each examinee. Although this score, called the pass/fail or cutoff score, is of importance to every examinee, it is also a focus of attention for parents, teachers, public interest groups, and other educators. Since the ramifications--legal, political, and financial--of setting the cutoff score are great, it is advisable to thoroughly consider the approach used.

Both the procedures and the issues discussed in this chapter are drawn from a study of minimum competency testing programs. While the procedures described are those that have been or are being used to set standards, the issues reflect a more general and comprehensive focus. At the state and local levels these issues surfaced in committee meetings and review sessions, at public gatherings and in print; not all were documented in state and local materials, many having been mentioned in the course of interviews. As a result, the issues represent a drawing together of many resources. Just as certain parameters may be taken into account in formu- lating competencies and testing instruments, so should they be considered when an approach to standard setting is selected. This chapter will high- light these parameters.

The standard setting strategies that will be discussed in this chapter are the following: (1) administrative decision or consensus, (2) Nedelsky, (3) Jaeger, and (4) contrasting groups. Appropriate examples of the application of each model will be presented. Since the situations, resources, and needs of each local district or state vary so much, however, no prescriptive rules will be presented. Rather, the procedures represent a subset of possible procedures, the issues a listing of those that the program planner may want to take into account in setting standards.

Issues and Parameters

Of the numerous procedures or strategies for setting the standards for a competency examination, some are brief and simple, while others are complex and time-consuming to implement. In the past, most major testing programs were norm-referenced and the cutoff score was usually established in relation to the strict statistical characteristics or outcomes of the test. With the exception of the Nedelsky method (1954), most procedures commonly in use now have been developed, tested, revised, and implemented in the past 15 years. And of these procedures, some have emerged directly as a result of needs arising from minimum competency testing programs.

Issues in the development, selection, and/or implementation of a standard setting strategy that are currently being considered by program managers include:

- legal defensibility
  -- legal issues
  -- uses of expert judgment
- ease of implementation
  -- time/expertise available
  -- reproducibility of procedures
- public acceptance
- psychometric characteristics
  -- single versus multiple cutoffs
  -- whether or not to include information about performance levels
  -- classification of examinee scores
- political considerations
- financial factors

The next section will discuss legal defensibility, implementation, and public acceptance. The section on strategies will present specific technical and psychometric characteristics of each model. Political and financial considerations will not be discussed.

Legal Defensibility

Legal issues. Although each of the above factors is important to consider in the standard setting process, preference and local need will dictate which issues will assume more or less importance. The legal defensibility of the test is one issue which deserves careful attention. Recently, in some notable cases, the courts have disallowed the use of a licensing or certification instrument because the cutoff score or passing score had been arbitrarily or capriciously established. In several decisions the courts have stated that although the required test standard or minimum performance level may be specified by the test user, such a score must bear a relationship to minimum job performance. In other decisions dealing with the statewide establishment and use of cutoff scores, the courts have ruled that for a standard to be valid and therefore appropriate for use, it must be job-related and logical (Dent v. West Virginia, 1899; U.S. v. State of North Carolina, 1975; U.S. v. State of South Carolina, 1977; Georgia Association of Educators v. Nix, 1975; Armstead v. Starkville Municipal Separate School District, 1975).

The legal consequences for misclassifying examinees on the basis of a minimum competency test may be very similar to those cited for certification and licensing tests. Setting standards by a well-documented, technically sound method will help to avoid those potential consequences.

Uses of judgment. Some years ago a major topic of discussion was whether judgment had a legitimate role to play in standard setting practices. In the course of the development and implementation of many different measurement procedures, specialists have come to recognize that varying amounts of judgment are employed in the establishment of any cutoff score (Shepard, 1979).

As a result, some researchers have revised their opinions. Jaeger, who initially classified standard setting models as either judgmental or empirical, currently holds the following view:

> All standard setting is judgmental. No amount of data collection, data analyses and model building can replace the ultimate judgmental act of deciding which levels of performance are meritorious or acceptable and which are unacceptable or inadequate. . . . In either case, subjective judgment of merit is inescapable (Jaeger, 1979, p. 48).

Strategies can vary with respect to the type of judgments they require and with respect to the factors that judges are asked to consider. In general, the extent to which the nature of the judgment can be controlled (so as to minimize extraneous influences) enhances the defensibility of a procedure.

As one example of a model which permits a standard to be set in an arbitrary and capricious fashion, Glass (1977) cites the example which he refers to as "counting backwards from 100%." In this model, the standard setters specify that 100% performance on each skill or objective is the desired outcome. In acknowledgment of a "certain" amount of human error, the required performance level is reduced from 100% to, say, 93% or 85%. What is arbitrary and capricious about this procedure and other equally unstructured approaches is the disregard for real factors and consequences on the part of those who set the standard.

For example, have the standard setters considered at what point in "counting backwards from 100%" ordinary human error can be confused with failure or noncompetency? At what point in the process is this issue considered or even identified? Is there any consideration for the issue of what percent of students will pass or fail as a result of one judge's estimate of error allowance against another's?

Many educators claim that this last factor (i.e., percent of students passing or failing) has little to do with competency assessment. Nonetheless, several other models do consider this particular factor as a means of facilitating a more focused judgment.

Glass points out that attempts to set standards are either "blatantly arbitrary" (as in the above example) or "derived from a set of arbitrary premises" (as in other, more structured models). Glass holds the view that the difficulty of setting standards well, however, does not excuse educators from doing so when needed; he goes on to caution: "Less arbitrariness is safer."

Psychometric Characteristics

The selection of a cutoff score has typically led program planners to consider the following issues:

-- whether to apply single or multiple cutoff scores;

-- whether information about examinee performance levels should be included or omitted;

-- whether the classification of examinee scores is correct.

Each of these issues will be discussed briefly.

Single versus multiple cutoff scores. Fundamental to this discussion of single versus multiple cutoff standards is an understanding of the difference between multiple (versus single) standards on a test which apply to all candidates and multiple standards which apply to different candidates. For an introductory discussion of the latter issue the reader is referred to Brickell (1977). This chapter will consider only the former issue: whether to establish single or multiple performance standards which every student must meet.

In setting the standards for a competency test, one should first consider the test purpose. If the purpose is to provide diagnostic information (instead of an overall descriptive determination), one will follow different avenues in setting standards. When the purpose of the test and its outcomes are clear, and these have been kept in mind throughout all the procedural and developmental steps of designing the competency test, the process of standard setting is facilitated. In order to decide in favor of either multiple cutoffs or a single cutoff to determine pass/fail decisions, the researcher who is considering the use of multiple cutoffs on a test may want to keep in mind the following points:

-- Requiring specified levels of performance on subtests or subsections within an exam ensures that every examinee classified as competent possesses some level of competence in each section of the domain.

-- If subtest cutoff scores are used, the stability of each subtest criterion will depend upon the number of items within each subsection.

-- One effect of establishing multiple cutoff scores (such that there are various criteria that must be met from subsection to subsection) is that the number of candidates who pass all sections will be reduced.

The cutoff score in each subsection need not be extremely high, since the aim is to ensure that examinees possess some level of skill on that subsection of the test. A trade-off is often necessary here, in order to keep the cutoff score from being unreasonably high on the subsections of the test, and yet still above the level of a chance score. In setting multiple cutoff scores, too, the possibility of misclassification is increased with each additional cutoff score. If there is only one cutoff score on the test, there is only one possibility for misclassifying candidates: at the point of the cutoff score. If there are four cutoff scores, one for each subsection of the test, there are four possibilities for misclassifying candidates. (See "Classification of Examinee Scores" for a discussion of issues of misclassification.)

If there is a single cutoff score, errors of misclassification may be reduced, but subskill performance is not ensured. For example, a candidate who achieves a high level of competency in one area of the test can in this way compensate for extremely low performance in another area.

In addition, one may choose to have a combination of both single and multiple cutoff score methods: that is, multiple cutoff scores for subsections of the test as well as a total cutoff score. Such a choice further increases the possibility of error, however. It will admit to the field of passing candidates only those who demonstrate a minimum level of competency in each of the subareas of the test and who, in addition, can meet some extra criterion in connection with the total test score; since no compensatory performance is allowed, it is likely that only a small number of examinees will pass.

Shepard (1979) indicates that the interpretation of data or the use of results from a test with multiple cutoff scores can be confounded in two ways: first, the cutoff scores may vary mostly as a function of variability in the judges' ratings and not as a function of differences in the importance or complexity of skills; second, the variability in difficulty of the test items used to measure domains for which there are different cutoff scores will affect the performance profile of examinees on those differently scored sections of an exam.

Airasian, Pedulla, and Madaus (1978) consider the decision for a single or multiple cutoff score in terms of the uses of testing results. While a total test score allows for pass/fail classification, it provides little information for diagnosis and remediation. This information is even less helpful when the test measures heterogeneous content. The authors indicate that there is no easy answer. The application of a single cutoff score may yield little diagnostic information, yet is clearly the easiest

method to administer. Multiple cutoffs, although they may increase classification error and increase administration and record keeping, generate more specific descriptive information about the individual examinee's competency.

Inclusion of examinee performance levels. As stated earlier, standard setting models involve reliance on expert judgment in setting a cutoff score. Even in the most structured models, judges rely on their educational experiences as students, teachers, administrators, etc., to help them set a benchmark or expected performance level. Another factor to consider is whether performance level (i.e., difficulty level or p-value) or other normative information should be included in the process of setting cutoff scores. Among programs which take into account this type of information about students in setting standards are Rocky River, Ohio and New Jersey.

Despite recommendations that item difficulty should influence the cutoff score (Klein and Kosecoff, 1973; Millman, 1974), several procedures to be considered involve ratings which are independent of actual examinee performance on the item. There are compelling arguments both for including and for excluding item difficulty in setting cutoff scores.

One issue that bears on this point is the purpose of the testing program. If one goal of the program is to classify students as masters or nonmasters on the basis of an "ideal" level of competency, this may decide the issue of whether or not the standard should be tied in any way to current examinee performance. When the cutoff score is to reflect an ideal level of competency which candidates must achieve, then current performance information is generally excluded.

In setting a cutoff score independent of normative data, the judges may define a standard that will result in the need for a great deal of improvement and change. On the other hand, the researcher may find that student performance is already quite close to the ideal level. When the cutoff score is set in relation to an ideal level of performance, educators can claim that the performance levels required for passing relate directly to a defined skill level which has been determined by experts. Such a standard may be said to be uncontaminated by information about current performance levels which might have led to a relaxation of the ideal standard.

A disadvantage to this approach is that the ideal level set by expert judges may bear little or no relationship to the current performance capabilities of students. The judges may conceptualize minimum competency in terms of experts, not in terms of the current examinees, and a large number of students may fail.

A modified version of an approach independent of performance is one in which overall test performance is provided as baseline data. (Typically, the inclusion of performance levels is accomplished on an item-by-item basis, or on the basis of a subarea or test section.)

Shepard (1979) claims that if judges create their own subjective models for normative data, there is a great risk that their comparisons will not be made on the basis of representative information. Therefore, she recommends providing representative data to the expert judges. Similarly, Conaway (1979) recommends that judges who set absolute standards for objective-referenced tests take the empirical difficulty of the items into account. He states that the effect of item difficulty on test scores is "pervasive." Therefore, judges who set standards for objective-referenced tests should have this information about item difficulty when reviewing the items which are the link between the objectives and the test scores. Very high or low standards might result if, indeed, these levels reflected only the judges' requirements. Normative information may facilitate the judges' task, since it can provide them with more guidance or focus. Judges may be able to incorporate into their decisions the factor of the percentage of students passing at various cutoff points (if that information is furnished) and review their ratings accordingly.

An inherent disadvantage in providing such normative information to the judges is that they may feel inhibited if their ratings depart from the empirical information provided. One may be concerned that standards set in this way merely mirror the status quo and provide no incentive for improving performance. Supplying normative information, such as percent of students passing at various cutoff scores, may mean that standards will be set largely on the basis of achieving a desired passing rate, without considering the content of the test or the level of competency deemed necessary relative to the domain measured. To provide too much information may defeat the original purpose of the task.

Classification of Examinee Scores

A critical issue in setting a standard is that of minimizing error or misclassification. This issue, in fact, usually takes precedence over all others in any discussion of standard setting, since it is the measure of success of the standard setting process. Florida's statewide assessment program, for example, is one which has paid particular attention to minimizing the risks of misclassification.

The explicit or implicit goal of a standard setting method is to achieve the maximally correct classification of examinees. If a student who is actually a master of the material being measured is classified as a nonmaster, the classification error is called "false-negative." Conversely, an actual nonmaster of the content who is assigned mastery status exemplifies a false-positive classification error.

To assess the extent to which a given standard setting methodology has accomplished the goal of classification, it is necessary to ask what proportion of the students tested have been rightly (or wrongly) classified. The answer to this question entails reference to some other "true" measure of a given student's mastery status. Where one is unlikely to obtain such a measure, one can appeal to some intuitive practice aimed at minimizing the proportion of students misclassified.

The problem is that it is difficult to minimize one type of classification error without affecting the likelihood of committing the other type of error. A standard set too low, for instance, is one that passes not only true masters but also some nonmasters. On the other hand, a standard set too high is one that fails not only true nonmasters, but also some masters.

Lowering the cutoff score reduces the likelihood of committing false-negative classification errors (because a larger proportion of the students will pass). Raising the cutoff score reduces the likelihood of false-positive classification errors (because a larger proportion of the students will fail).

Therein lies the problem. What is the optimal point at which the standard should be set so that an appropriate compromise is made between the two types of errors? Should the trade-off between false-positive and false-negative errors be decided in favor of one or the other; or should one favor neither, and instead seek an evenly balanced compromise? Asking how serious each type of error is when placed in the context of the purpose of the test or the use of test results is one way of answering these questions.

In practice, then, the question can be answered in terms of the probable effects of committing the two different types of errors. It is clear that to generate decisions of either the false-positive or false-negative type could have serious implications for individual students, for teachers and administrators, for policy-makers, and for testing programs as a whole.

The implications of false-negative decisions include: (1) the psychological and social burden to be borne by students who are incorrectly classified as nonmasters; (2) the culpability (either ethical or legal) of

the decision maker in such an instance; (3) the costs (both in dollars and human resources) of providing remediation to students who do not, in fact, require it; (4) the costs of retaining students in grade in larger proportions than had previously been expected or encountered; and (5) loss of confidence in the validity of the test instruments and in the decisions emerging from their administration.

The implications of false-positive decisions include: (1) unfounded aspirations for success in competency-based endeavors on the part of students wrongly classified as masters, (2) unfounded expectations on the part of potential employers about the skill levels of such students as potential employees, and (3) the perception of the lay public that graduates lack sufficient command of skills (leading to a loss of confidence in the value of the high school diploma).

The purpose of the tests or testing program and the use made of the results will help to determine the seriousness of one type of error or the other.

The ability of a particular method to classify examinees correctly is a prerequisite for selecting that standard setting approach. It is, however, one that should be viewed in terms of the degree or extent to which correct classification is maximized. There is no procedure known that will correctly classify 100% of the examinees. Many cutoff score models approach this issue in different ways. In some methods, such as decision-theoretic approaches (Hambleton & Novick, 1973), an explicit emphasis is placed on controlling the amount of misclassification. In models which involve judgments on test questions (Nedelsky, 1954), the control is implicit in correct application of the model.

Airasian et al. (1978) also raise the issue of classification accuracy as it relates to public acceptability. Suppose that a method of setting a cutoff score maximizes correct classification, but has the result that an overwhelming percentage of the examinees fail the test? Such a method risks failure not because of its statistical weakness, but because to pass so few students will have enormous and far-reaching educational, psychological, and financial consequences.

Ease of Implementation

The establishment of cutoff scores generally requires not only statistical sophistication, but also an awareness of certain political, educational, and financial concerns. In either selecting a cutoff score model

which has been used previously or creating an approach tailored to the needs of a particular situation, therefore, program planners may want to consider the following factors:

-- time and technical expertise: what is available versus what is needed;

-- reproducibility of procedures.

The issue of public acceptance will be discussed separately.

Time and technical expertise. The developmental phases of a minimum competency program can be quite lengthy, particularly when they involve determination, definition, and resolution of complex political and theoretical issues. While exercising care in the determination of a cutoff score is generally desirable, the task need not be so time-consuming that it hinders the completion of the project. Procedures that require input from large numbers of people for a substantial amount of time are cumbersome to implement and very costly in terms of professional time (e.g., Jaeger, 1978). Moreover, there are some procedures which require judges to make several ratings or judgments. In such a process, the judges can become confused or the scores collected can be unreliable because of the complexity of the required task.

Reproducibility of procedures. A feature of some competency tests is that they measure skills in the context of "real life" situations. To the extent that this is true, the timeliness and appropriateness of test items may be of concern and may therefore need to be reviewed periodically. Even in programs in which the assessment instrument is not written in life role terms, the test items measure objectives, the importance of which may vary over time. Whenever the objectives and/or items change, the cutoff score or standard may be affected.

In some testing programs, a desire for test security has entailed the generation of multiple forms of a test. In these cases and others in which new tests are developed and introduced frequently, it may be necessary to recalculate or reapply a cutoff score method to a new set of test questions. In these cases, a standard setting procedure that is not unduly complex or expensive, but still sound, has been found to be most useful.

Public Acceptance

On the basis of evidence from the program in Kanawha County, West Virginia, Candor-Chandler (1978) states that a primary consideration in the implementation of a minimum competency testing program should be whether the model can be easily understood by the community. Public acceptance is likely to be facilitated if the appoaches taken in developing and implementing the program are understandable and acceptable to teachers, administrators, and the wider community. As a key component of the program, standard setting may merit particular consideration with respect to the issue of public acceptance.

One notable outcome of a competency program is the number or percent of examinees who pass or fail. According to the example from Airasian et al. (1978) was cited earlier, although a standard setting approach may statistically maximize the correct classification of examinees, it may fail more students than the constituencies of a program find acceptable. Standard setters often take this factor into account. They may estimate the minimum percentage of examinee failure that the public will tolerate; or they may set a cutoff score to achieve a specific percent of passes and fails.

Miller (1978), in reporting on national conferences on minimum competency testing, states that it is the process for selecting the cutoff score which is the key to its acceptability. Both community representatives and experts in the field can contribute important information to the process. Furthermore, it is recommended that the standard setting process not be viewed as a single or isolated task, but rather as one that should be reviewed from time to time by different judges, revised on the basis of field data, and reconsidered in the light of possible changes in the goals or emphasis of a particular minimum competency program (Fremer, 1977; Miller, 1978; Shepard, 1979).

Standard Setting Strategies

In the following discussion of standard setting models currently in use in the field, highly empirical models will be excluded. It has been found that highly statistical models are not feasible in actual practice because they require conditions which cannot be met. For example, Millman

(1973) has defined a model for use only with individual students. The Emrick model (1971) requires homogeneity, an equal level of item difficulty, and equal item intercorrelations. Other Bayesian approaches require collateral and prior information (Hambleton & Novick, 1973) which is often difficult to obtain. All of these empirical models also incorporate judgments (Block, 1972; Millman, 1973).

The models that will be discussed are:

(1) administrative decision or consensus;
(2) Nedelsky;
(3) Jaeger;
(4) contrasting groups;

With the exception of administrative decision or consensus, the methods above can be classified as requiring either (1) judgments on items or (2) judgments on examinees. This distinction is also used in part by Hambleton and Eignor (1978) and Zieky and Livingston (1977), and the reader is referred to these works for additional discussion.

Administrative Decision or Consensus

Neither the administrative decision nor the consensus method of setting cutoff scores can be classified on the dimensions of judgment or of statistical assumptions, because there is very little structure or dimensionality to analyze in either approach. They are included in this discussion because, for a variety of reasons, they are the methods which are most commonly employed.

Setting standards by administrative decision means simply that the cutoff score is determined by one or more persons holding a position of authority or responsibility in a testing program. Although these judges may be capable of making an extremely informed decision, it may not be a decision which is open to external verification of its appropriateness. As a result, a disproportionate number of students may pass or fail the test because, in setting the standard, there was no accommodation for the pass/fail rate.

The second and very similar method for establishing a cutoff score is by consensus. The procedure for setting the cutoff score may again be largely undefined, but the judges in this method are usually members of a

group which is large enough to minimize the outlook of any one individual. Also, such a group usually consists of educators representing various educational constituencies, so that a complete array of educational beliefs is brought to the issue of setting a passing score (Wilson, 1976).

Standard setting by administrative decision or by consensus is popular for a great many reasons. As a first effort toward standard setting, these approaches are easy for all of the participants in a program to understand. They are not time-consuming or costly methods, and require no additional technical expertise. What these two procedures may lack in statistical strength they compensate for in other areas. For example, they accommodate certain issues better than many other models. Financial, political, and public concerns weigh very heavily and are usually carefully considered in these standard setting processes. The judges involved are often acutely aware of the importance of these issues.

It should be noted that one aspect of the consensus method, that of group decisions or recommendations by expert judges, is a major component of many of the procedures which will be described below. Each of the other procedures, however, includes structured review requirements and/or empirical information.

Setting standards by administrative decision or consensus may also involve considering the specific competencies to be assessed. In Vermont, for example, administrators prepared a list of competencies in five areas following statewide reviews and for each competency set an individual standard. Standards below 100% were set only for those competencies on which a student might make errors due to carelessness rather than lack of mastery. Such competencies are those measuring processes (e.g., writing names of arabic numerals) rather than a student's command of facts. Where processes are being assessed, the Department defines 80% as meaning that the pupil must answer correctly at least 80% of the examples.

Administrators responsible for setting standards may also consider using field-test data in arriving at a decision. In Maryland, for example, Project Basic staff reviewed the results of a field test of four reading competencies before setting a passing standard of 80% on each competency for the secondary-level Functional Reading Test. For a more complete listing of state and local programs using field-test data and specific standard-setting procedures, see A Study of Minimum Competency Testing Programs: Final Summary and Analysis Report (National Evaluation Systems, 1979).

Judgments on Items

The methods to be described here are the Nedelsky approach, and the model proposed by Jaeger. These are methods which require specialists to examine a test or its items and to decide on the score which a person with minimum competency should attain.

Nedelsky

One of the most popular approaches for setting standards for minimum competency programs is one that was originally developed for use on examinations in medicine. The Nedelsky approach is flexible enough for use on any number of test items--i.e., a test of any length. The ratings can be completed with or without normative data. The number of judges or raters can vary. Nedelsky's approach can be used only on multiple-choice items, for which there is a single correct response.

Glass (1978) has outlined the Nedelsky procedure as follows:

> Directions to Instructors
>
> Before the test is given, the instructors in the course are given copies of the test, and the following directions:
>
> In each item of the test, cross out those responses which the lowest D-student should be able to reject as incorrect. To the left of the item, write the reciprocal of the number of the remaining responses. Thus if you cross out one out of five responses, write 1/4.
>
> Example. (The example should preferably be one of the items of the test in question.)

Light has wave characteristics. Which of the following is the best experimental evidence for this statement?

A Light can be reflected by a mirror.
B Light forms dark and light bands on passing through a small opening.
C A beam of white light can be broken into its component colors by a prism.
1/4 D Light carries energy.
E Light operates a photoelectric cell.

Preliminary Agreement on Standards

After the instructors have marked some five or six items following the directions above, it is recommended that they hold a brief conference to compare and discuss the standards they have used. It may also be well that at this time they agree on a tentative value of constant k (see section on The Minimum Passing Score). After such a conference the instructors should proceed independently.

Terminology

In describing the method of computing the score corresponding to the lowest D the following terminology is convenient:

a. Responses which the lowest D-student should be able to reject as incorrect, and which therefore should be primarily attractive to F-students, are called F-responses. In the example above, response E was the only F-response in the opinion of the instructor who marked the item.

b. Students who possess just enough knowledge to reject F-responses and must choose among the remaining responses at random are called F-D students, to suggest borderline knowledge between F and D.

c. The most probable mean score of the F-D students on a test is called the F-D guess score and is denoted by M_{FD}. As will be shown later, M_{FD} is equal to the sum of the reciprocals of the numbers of responses other than F-responses. (In the example above, the reciprocal is 1/4.)

d. The most probable value of the standard deviation corresponding to M_{FD} is denoted by $_{FD}$.

It should be clear that "F-D students" is a statistical abstraction. The student who can reject the F-responses for every item of a test and yet will choose at random among the rest of the responses probably does not exist; rather, scores equal to M_{FD} will be obtained by students whose patterns of responses vary widely.

The Minimum Passing Score

The score corresponding to the lowest D is set equal to $\bar{M}_{FD} + k\ _{FD}$, where $\bar{M}_{FD}$ is the mean of the M_{FD} obtained by various instructors, and k is a constant whose value is determined by several considerations. The F-D students are characterized not so much by the positive knowledge they possess as by being able to avoid certain misjudgments. Most instructors who have used the F-D guess score technique have felt that this "absence of ignorance" standard is a mild one, and that therefore the minimum passing score should be such as to fail the majority of F-D students. Assigning to k the values -1, 0, 1, and 2 will (on the average) fail respectively 16 percent, 50 percent, 84 percent, and 98 percent of the F-D students. An informed final decision on the value of k can be reached

after the instructors have chosen the F-responses, for at that time they are in a better position to estimate the rigor of the standards they have been using. In keeping within the spirit of absolute standards, however, the value of k should be agreed on before the values of M_{FD} are computed and certainly before the students' scores are known.

It is the essence of the proposed technique that the standard of achievement is arrived at by a detailed consideration of individual items of the test. Only minor adjustments should be effected by varying the value of k. The reason for introducing constant k, with the attendant flexibility and ambiguity, is that F-responses in most examinations vary between two extremes: the very wrong, the choice of which indicates gross ignorance, and the moderately wrong, the rejection of which indicates passing knowledge. If a particular test has predominantly the first kind of F-responses, this peculiarity of the test can be corrected for by giving k a high value. Similarly, a low value of k will correct for the predominance of the second kind of F-responses. It is expected that in the majority of cases a change of not more than $\pm$ .5 in the tentative value of k agreed upon during the preliminary conference should introduce the necessary correction. It would be difficult to find a theoretical justification for values of k as high as two; for more tests the value k = 0 is probably too low. This suggests a rather narrow working range of values, say between .5 and 1.5 with the value k = 1 as a good starting point.

If a part A of a given test consists of N_A items, each of which has s_A non F-responses (one of these being the right response), the F-D guess score for each item, i.e., the probability that an F-D student will get the right answer in any one item, is $p_A = 1/s_A$. The most probable values of the mean and the square of the standard deviation on this part of the test are given by $M_A = p_A N_A$ and

$_A = p_A(1 - p_A)N_A$. M_{FD} and $_{FD} = _A _A$. The value of M_{FD} must be accurately computed for each test. $_{FD}$, however, may be given an approximate valuie. In a test of five-response items s may vary from one to five. If these five values are equally frequent, $_{FD} = .41$ N. If, on the other hand, the extreme values, s = 1 and s = 5, are less frequent than the other three values, as seems likely to be true for most tests, .41 N $_{FD}$ .50 N. Since k $_{FD}$ is usually much smaller than M_{FD}, approximations are in order. With
With k = 1 and $_{FD} = .45$ N, the equation, Minimum Passing Score = $\bar{M}_{FD}$ + .45 N, should work out fairly well in the majority of cases and is therefore recommended as a starting point in experimenting with the proposed technique (Glass, 1978, pp. 22-24).

Adaptation/application. Since minimum competency testing programs specify a standard not in terms of traditional D or F classroom scores, Nedelsky's procedure has been adapted in a number of ways. Nedelsky's procedure is also often coupled with the Angoff method, although the latter is more typically used in setting standards for licensing examinations. New Jersey, in its minimum Basic Skills program, used both methods, but modified Nedelsky's procedure in the following way:

1) The first step in applying the standard setting procedure is to think about what you consider to be the lowest level of performance you are still willing to classify as mastery of the skills measured by the test that you worked on. If you have recent classroom experience, it may help you to think about students you have known that were just barely good enough to be considered masters of the basic skills measured by the test.

We expect that there will be some differences of opinion as to what is meant by minimally acceptable performance.

2) The second step is to look at the first question in the test and decide how many wrong answers are so wrong that even the minimally acceptable student would know that they are wrong.

For example, the following question is similar to one on the Grade Three Math test:

> The school lunchroom served 506 people on Monday and 315 people on Tuesday. How many people were served on the two days?
>
> (A) 191
> (B) 201
> (C) 811
> (D) 821

You may decide that even the minimally competent student should know that A and B are wrong because the total for two days would be greater than the number on any single day. But you may decide that wrong answer C involves an error that the minimally competent student would not know is wrong. You would therefore decide that two wrong answers for the questions are so wrong that even the minimally competent students would know that they are wrong.

3) We will then ask for a few volunteers to tell the groups which wrong answers were selected and their reasons for selecting them. You will be encouraged to discuss the choices. The discussions may either confirm your earlier opinions or change your mind.

4) The last step is for you to record the number of wrong answers you selected as being so wrong that even the minimally qualified student would know they are wrong.

5) We will go on to the next question and repeat the process. After you are done, we will estimate the tentative standard for each test based on the data you provided.

The committees utilized the modified Nedelsky procedure and each person developed an estimated proficiency standard for a particular test. Next, a mean estimated standard was obtained. This mean was the best estimate for the proficiency standard using the Nedelsky procedure (Koffler, 1979, pp. 9-10).

Application. The Nedelsky model was applied by the Kanawha County schools in West Virginia (Candor-Chandler, 1978). Although consistency was found across groups of judges who completed the process at three different times, the researcher reports that the application was not successful. Teachers were uncomfortable with the process of setting standards and of determining minimum competency. In addition, it was found that teachers tended to estimate that there was less probability that students would get "easier" items right than the more "difficult" items. The judges felt that a student could get an easy item wrong by making a simple error, but in the case of the more difficult items, "the students either knew it or they didn't."

Candor-Chandler indicated that the cutoff scores for Kanawha County were then set after a review of preliminary data and consideration of certain educational/instructional factors.

C295

Jaeger

The model proposed by Jaeger (1978):

(1) is technically straightforward, quite long, and maximizes participation and involvement of educational constituencies;

(2) is an iterative process;

(3) involves normative data in part of the review.

It should be noted that this model, unlike some others, defines minimum competency without using that term in the body of the definition, and so avoids circularity.

Jaeger proposed this method for standard setting for the North Carolina high school competency test. To accomplish this task, 700 persons (registered voters, teachers, counselors, and administrators) convened in groups of 50 to proceed through the standard setting model.

Judges were first required to take the exam which they would later rate. For each item judges were asked one of the following two questions:

(1) Should every high school graduate be able to answer this item correctly?

(2) If a student does not answer this item correctly, should s/he be denied a high school diploma?

Judges next received the results of the above survey questions as well as actual performance data. With this information, judges were asked to review and revise their initial judgments as they might consider necessary.

Jaeger's procedure calls for recalculation of the judges' ratings, redistribution of the new ratings, and another judgment. Judges then received information on the proportion of students who would have passed or failed, as determined on the basis of the recommended cutoff scores. With this information, judges were asked to make a final statement on the "necessity" for each item on the test.

Median scores were calculated by group (type or constituency), and the passing score was then set at the minimum median score calculated for a group.

Adaptation/application. The Gallagher report to the North Carolina Board of Education (1978) stated that there was a delay in setting standards until the completion of four studies, designed to provide additional decision-making information. The studies consisted of:

(1) a comparison of competency test results with norm-referenced test results;

(2) identification of the minimally competent and incompetent student;

(3) teacher judgment of the tests;

(4) a statistical study of the spring (1978) trial distributions.

In support of (1), scores from the SHARP Reading and TOPICS Mathematics were compared to the California Achievement Test. Both the total score and the separate reading and math scores were reviewed for the total group tested and for subgroups classified by sex and race. "All of the results support the need to place (these) raw scores or percentage scores into some more standardized set of measures that would allow one to make some legitimate comparisons across subject areas" (North Carolina, SDE, 1978, p. 11).

For the second study, schools in a sample group were asked to identify students whom they considered marginally competent and students considered noncompetent. The performances of these students on the various tests led the author to stress the need for differentiated cutoff scores in different subject areas.

The procedure used for the third study is very similar to that proposed in Jaeger (1978). Specifically, teachers and other curriculum specialists participated in a one-day conference for the purpose of giving judgments as to a minimum passing score for North Carolina on the SHARP and TOPICS tests. The judges' tasks were to:

(1) take the test and try to see the test through the eyes of a competent (not superior) student;

(2) judge the percent of correct answers that should be required as passing scores for the reading and mathematics tests;

(3) review and revise their original judgments as necessary, when given student trial performance data (it is interesting to note that the math standard was reduced, while the reading standard was relatively unchanged as a result of this step);

(4) review and revise the second judgment made, if necessary, when given the group results on the recommended standard.

Gallagher notes that the ratings which teachers made for the math test changed with the increased information provided to them at each step. The teachers believed that the information provided assisted them in making informed judgments.

The fourth study was a focused statistical analysis of the number and placement of items students omitted from their responses. Time and/or motivation seemed to be relevant factors in accounting for the increased number of items omitted in the last part of the test.

With all of the information from the four studies, the North Carolina Competency Test Commission met and established the standards for the reading and math tests.

Judgments on Examinees

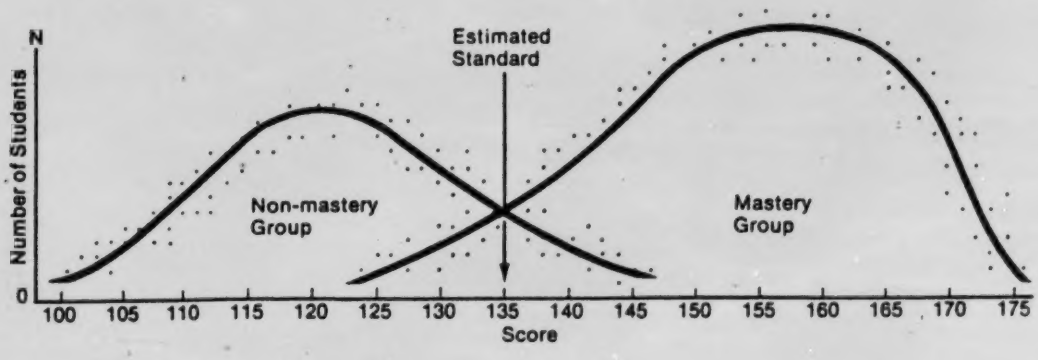
Two methods for setting cutoff scores proposed by Zieky and Livingston (1977) respond directly to many concerns encountered in minimum competency assessment. These methods, called the "borderline groups" and "contrasting groups" methods, require judges to make judgments on examinees, and not on the test or its items.

Contrasting Groups

As the name implies, the contrasting groups method involves examination of scores of students classified in discrete groups: those considered to be masters of the material measured by the test (for which the standard is to be set) and those considered to be nonmasters.

Judges who are familiar with each student's current capabilities in the content of the test are asked to identify those students who are clearly masters and those who are clearly nonmasters. According to Zieky and Livingston (1977), a minimum of 100 classified students is needed to achieve a stable estimate of the standard.

Following the test administration, the score distributions of the students in these two distinct groups are superimposed on each other. An initial standard for the test is the intersection point of the two graphs. An advantage of this method is that the cutoff score can be adjusted (raised or lowered) to minimize a selected error of classification. The following table illustrates this method.*



In this method, the graphic representation of score distributions facilitates the consideration of errors of misclassification. While it is

---

* From Manual for Setting Standards on the Basic Skills Assessment Tests, by M. Zieky and S. Livingston. Princeton, New Jersey: Educational Testing Service, 1977.

possible in other models to recalculate percents of students passing or failing by adjustments to the standards, some researchers may prefer the visual presentation--an integral part of the contrasting groups approach.

Application. A procedure used by Fillbrandt and Merz (1977) to set standards for a California school district is similar in concept to the contrasting groups approach of Zieky and Livingston (1977) and the optimal cutting score method of Berk (1976). The researchers determined that to distinguish between students who are competent and noncompetent, they would test and establish standards on the basis of the performance of "successful" persons in the community. Fillbrandt and Merz used matrix (test item and examinee) sampling to minimize the test-taking time of participants selected as meeting the criteria specified for "successfully employed persons."

Standards were set on the basis of the empirical results of the test. For example,

Score Distributions Derived from Multiple Matrix Sampling

Parameters	Reading Test	Math Test
Mean	25.63	29.88
Standard Deviation	4.63	9.80
Median	27.14	31.00
Q + 2.09	7.27	
90th %ile	30.00	42.00
75th %ile	29.00	38.00
50th %ile	27.00	31.00
25th %ile	23.00	23.00
10th %ile	19.00	16.00
Reliability	.854	.916

A cutting score of 20 was established for the reading test. This decision was based on plots of the distribution which indicated that an asymptote was reached near the scores of 19 and 20; it appeared that below the score of 19 the curve flattened, indicating that the percentages of those scoring at each point below 20 were about equal. In addition, the score of 20 represents 66.6% correct and identifies the upper 90% of scores (Fillbrandt & Merz, 1977).

Two other programs that have used this standard-setting procedure are Kentucky and Peterborough, New Hampshire. In Kentucky, the Department of Education asked a representative sample of teachers to classify their students into three groups: those who do or do not need remediation; and those who may need remediation in the specific competencies. The students then took the screening test on which a standard was to be set. The standard chosen was the point of intersection between the scores of students who do need remediation and those who may need it. In Peterborough, New Hampshire, a standard for each competency in communication and computation was set by comparing the scores of students two grade levels ahead and two behind the grade level at which mastery of the competency is expected.

The success of this procedure has been attributed to the involvement of the community and the definition of standards in terms of functional competencies actually needed in the job market; in addition, the complexity and technical detail of the study furnish very strong evidence for its acceptability.

Wilson (1976) describes the use of an external criterion group, such as the one utilized here, as a better approach to standard setting than administrative decision or consensus. He also acknowledges that to use such a group is more expensive and more difficult in terms of the technical expertise and logistics which are necessary.

Program planners may also wish to consider a companion procedure to the contrasting groups method known as the borderline groups. In this method at least 100 students whose performance cannot be clearly classified as adequate or inadequate are tested. The median of the scores of this group is computed and used as an estimate of the cutoff score.

* * * * * * * * * *

Whichever method is used, the ease of implementation is enhanced by the use of procedures that are simple yet sound, and neither costly nor time-consuming; both are based on the judgments of teachers who are extremely knowledgeable about student capabilities. This last factor can present a problem if teachers are not carefully selected and trained and if their judgments are not accurate with respect to the classification of specific students. These two approaches also rely on a definition of minimum competency relative to the content being tested, and not one directly related to the test. This definition of minimum competency must be applied to classify students into groups for the statistical analyses required by the models. It is therefore critical to the accuracy of the tests.

What Is Actually Being Done

This chapter has cited examples of the application or modified application of each of the standard setting procedures selected for discussion. These examples have been drawn from descriptions of both state and local minimum competency programs. In addition, the following table summarizes information about the total number of programs which employ each of the various procedures to set their standards.

Procedures Used in Setting Standards*

Procedure	State	Local
Administrative Decision	5	6
Contrasting Groups	2	3
Nedelsky/Angoff	1	2
Field Test Results and/or Other Statistical Procedures	9	7
Competency Definition	3	2

* From National Evaluation Systems, 1979. The reader is referred to this report for additional information.

In the table above, the procedure labeled Competency Definition is a process in which the standard is established as part of the competency definition. The procedure or method for this is not specified.

Similarly, several states specify standards using field test data and/or statistical techniques. This in itself is unlikely to be the procedure, but only a material adjunct to a process such as administrative decision, consensus, or Nedelsky. In addition, there are very few statistical techniques that generate a standard. Again, most represent a component of the process. Further information or details which would tie the techniques to a procedure were not available.

References

Airasian, P., Pedulla, J., & Madaus, G. Policy issues in minimal competency testing and a comparison of implementation models. Boston: Heuristics, 1978.

Angoff, W. H. Scales, norms, and equivalent scores. In. R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Armstead v. Starkville Municipal Separate School District, 325 F. Supp. 560 (N.D. Miss. 1971), Modified, 461 F. 2d. 276 (1972).

Berk, R. A. Determination of optimal cutting scores in criterion-referenced measurement. Journal of Experimental Education, 1976, 45, 4-9.

Block, J. H. Student learning and the setting of mastery performance standards. Educational Horizons, 1972, 50, 183-190.

Brickell, H. M. Seven key notes on competency testing. In B. S. Miller (Ed.), Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

Candor-Chandler, C. Competency measurement at the local level: A case study of Kanawha County Schools, West Virginia. In R. B. Ingle, M. R. Carroll, & W. J. Gephart (Eds.), The assessment of student competence in the public schools. Bloomington, Indiana: Phi Delta Kappa, 1978.

Conaway, L. E. Setting standards in competency-based education: Some current practices and concerns. In M. A. Bunda & J. Sanders (Eds.), Practices and problems in competency-based measurement. NCME, 1979.

Dent v. West Virginia, 129 U.S. 114 (1889).

Ebel, R. L. Essentials of educational measurement. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

Fillbrandt, J. R., & Merz, W. R. The assessment of competency in reading and mathematics using community-based standards. Educational Research Quarterly, 1977, 2(1).

Fremer, J. Setting and evaluating competency standards for awarding high school diplomas. Paper presented at the meeting of the National Council on Measurement in Education, New York, April 1977.

Georgia Association of Educators v. Nix, 407 F. Supp. 1102 (1976).

Glass, G. Standards and criteria. Journal of Educational Measurement, 1978, 15, 237-261.

Hambleton, R. K., & Eignor, D. R. Competency test development, validation, and standard-setting (Research Rep. No. 84, Laboratory of Psychometric and Evaluative Research). Amherst, Massachusetts: University of Massachusetts, School of Education, 1978.

Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Jaeger, R. M. A proposal for setting a standard on the North Carolina High School Competency Test. Paper presented at the meeting of the North Carolina Association for Research in Education, Chapel Hill, North Carolina, 1978.

Jaeger, R. M. Measurement consequences of selected standard-setting models. In M. A. Bunda & J. Sanders (Eds.), Practices and problems in competency-based measurement. NCME, 1979.

Klein, S. B., & Kosecoff, J. Issues and procedures in the development of criterion-referenced tests. Princeton, New Jersey, 1973. (ERIC Document Reproduction Service No. TM26)

Koffler, S. L. Setting proficiency standards: A comparative approach. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Miller, B. S. (Ed.) Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 43(2), 205-216.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, California: McCutchan Publishing, 1974.

Nassif, P. M. Standard-setting for criterion-referenced teacher licensing tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, 1978.

Nedelsky, L. Absolute grading standards for objective tests. Educational and Psychological Measurement, 1954, 14, 3-19.

North Carolina, State Department of Education. Setting minimum competency standards (Report of the North Carolina Competency Test Commission, J. J. Gallagher, Chairman). September 1978.

Shepard, L. A. Setting standards. In M. A. Bunda & J. Sanders (Eds.), Practices and problems in competency-based measurement. NCME, 1979.

United States v. North Carolina, 400 F. Supp. 343 (E.D.N.C. 1975), 425 F. Supp. 789 (E.D.S.C. 1977).

United States v. South Carolina, 15 FEP Cases 1196 (D.C.S.C. 1977).

Wilson, H. A. Two sides to tests: Positive, negative. National Assessment, 1976.

Zieky, M. Steps to follow and questions to consider when setting standards. Handout prepared for symposium on Critical issues in setting minimum competency standards, Spring 1979.

Zieky, M., & Livingston, S. Manual for setting standards on the basic skills assessment tests. Princeton, New Jersey: Educational Testing Service, 1977.

CHAPTER 5

INTEGRATING TESTING WITH INSTRUCTION

Mary F. Tobin

Introduction

The rise of minimum competency testing has spurred renewed interest in curricular and instructional issues, ranging from speculations about the impact of such testing upon the curriculum to discussions of how test results can be used most effectively. Some observers, for example, fear that the implementation of minimum competency testing programs will lead to a narrowing of the curriculum, while others have speculated that an increased focus upon test results will undermine credence in the professional judgments of teachers.

Nonetheless, both critics and proponents of minimum competency testing suggest that a significant challenge which administrators and program planners face in implementing a minimum competency program is to develop a course of instruction for students who will take such tests in the future, as well as for students who have already failed the tests. As Ryan (1979) and Shoemaker (1979) point out, a testing program will neither improve nor guarantee learning. An additional problem that confronts those planning a minimum competency testing program is to develop testing activities that are an integral part of the instructional program. The purpose of this chapter is to discuss how different programs have resolved these issues and to summarize the suggestions and comments of program planners.

A fundamental assumption of this chapter is that integrating testing with instruction means ensuring that testing activities provide appropriate information to the personnel responsible for decisions that affect students, the curriculum, and instruction. Those responsible for these decisions can include, for example, classroom teachers, the school or district curriculum coordinator, school or district administrators, and state-level personnel (e.g., Department of Education staff). Just as the nature of their decisions vary, so will their information needs differ.

The first step to help ensure that the testing activities provide useful information is to identify who will use the test results and for what purposes. This chapter is specifically concerned with those uses

relating to curriculum and instruction and also with those groups of people who typically use test results in making decisions related to curriculum and instruction. The first part of this chapter will present examples of the ways in which teachers, local curriculum coordinators, administrators, and state-level personnel use minimum competency test results in altering and assessing curriculum and instruction.

In determining to what extent test results will be used in making instructional and curricular decisions, administrators and planners might want to consider ways to promote the use of test results by key consumers (e.g., teachers, school personnel). Some programs have developed methods for encouraging the use of test results. These will also be described.

Program planners responsible for developing instructional programs both to introduce the competencies to students and to provide remediation may wish to consider a variety of organizational arrangements. Programs that have been implemented yield examples of possible arrangements. Some options may be more appropriate to providing an introduction to the competencies rather than remediation, and vice versa. Factors to consider in choosing one arrangement over another include the number of students, the size of the instructional staff, the availability of curriculum materials related to the competencies, the physical facilities, the ability and interest on the part of the staff in providing remediation, and the possibility of using paraprofessionals and volunteers. The second part of this chapter will discuss possible arrangements and how these factors influence the choice of options.

In the third part of this chapter the general issue of how to integrate testing and instruction is explored from a more comprehensive perspective. This discussion will treat the development of program components and the consequences for the instructional program. For example, the choice of the testing schedule can have potential consequences for the instructional program, such as ensuring that the staff which is to provide remediation have the necessary training to do so. This discussion is intended to point out that in dealing with the issue of how to integrate testing with instruction, the methods chosen need not be limited merely to using test results in more and better ways. Rather, a more comprehensive view may be taken in which the testing and the instructional programs are designed to complement each other.

The discussion below is therefore intended to provide a general introduction to the following topics: the key audiences who might use test results in making decisions affecting curriculum and instruction, ways to encourage the use of test results by members of these groups, options for organizing regular and remedial instruction in the competencies, and suggestions for integrating testing with instruction based on

the design of specific program components. Where possible, ongoing programs will be used to illustrate the options available to program planners. The discussion does not assume that program planners have identified how test results will be used, or that they have decided to alter the curricula. Rather, the ways in which test results are typically used are described, and general suggestions for organizing instruction in the competencies and for integrating various components of the testing program with the instructional program are included.

MCT Results and Decisions Related to Curriculum and Instruction

As noted above, key audiences who might use test results in making decisions concerning curriculum and instruction include teachers, school or local district curriculum coordinators, local administrators, and state-level personnel (e.g., legislators, Department of Education staff members). Test results may be used for diagnostic purposes in working with individual students, for assessing the strengths and weaknesses of a particular course or program, or for assessing the strengths and weaknesses of a school district's instructional and curricular program. In most cases, whatever the purpose, test results are used in conjunction with other information. A teacher might review a student's records in reading as well as the results of a minimum competency test in order to determine whether the student required remediation. In few instances do administrators and planners consider that minimum competency testing yields all the information needed to make specific decisions.

Using Test Results for Diagnostic Purposes

Hillsborough County, Florida, offers one example of a program in which both test results from the statewide assessments administered in grades 8 and 12 as well as the results of locally developed minimum competency tests are used to identify students in need of remediation. Administrators in this program have developed a compensatory education program for grades 7-12. Once identified, students are assigned to special classes in which diagnostic tests are first given to determine specific areas of weakness.

A posttest is also administered in these classes to measure students' progress and these results are used to determine whether more remediation is required.

The extent to which the results from minimum competency tests can yield diagnostic information is a subject of debate among educators since these tests typically indicate only whether a student has or has not mastered the competencies. Some have recommended that test results be used primarily for screening to identify students who require remediation, and that results be used in conjunction with other indicators, such as teacher judgments. As Means points out, "If a student fails a test of minimal competency in reading comprehension, the presumption underlying the model is that diagnosis of the reading comprehension problems must be successfully completed and that inferences must be made about the diagnostic test data so that instruction can be prescribed. Yet, the task of diagnosing problems related to reading comprehension is difficult because at present test makers cannot factor discrete reading skills out of the tests" (Means, 1979, p. 5).

Means goes on to suggest that even given the absence of diagnostic information from current popular tests of reading comprehension, "the talented reading teacher may be able to successfully prescribe instruction in reading comprehension" on the basis of test results which merely indicate general problems in this area. Hence, one issue administrators and planners may face in determining how test results will be used and in developing minimum competency tests is the extent to which testing both can and will be used to yield diagnostic information. As in the case of the Hillsborough County program, one option is to use results to identify students requiring remediation, and then to obtain diagnostic information useful in prescribing instruction through other means (e.g., other testing, consultation with teachers trained in diagnostic techniques).

Ways to encourage appropriate uses of test results among teachers include workshops and staff meetings in which the uses and limits of the test data are discussed. The role of teacher judgments vis-a-vis the test data is an issue administrators may wish to consider carefully. In some programs (e.g., Fitchburg, Massachusetts) the teacher uses test results in conjunction with a personal judgment to assess the progress of students in basic skill areas. Too great a reliance upon test results may lead to the neglect of other useful information about a student's learning difficulties and their causes; minimizing the role of testing, however, may result in the program being perceived as a pointless demand upon staff time. Given estimates of the extent to which standardized test results are used by teachers (see Goslin, Epstein, & Hallock, 1979), administrators who perceive minimum competency testing as yielding useful information may be interested in trying a variety of procedures for encouraging the staff to

use test data. Administrators may also want to consider instituting such procedures as periodic surveys or interviews to uncover particular obstacles (e.g., obscure report forms, lack of interest, or hostility) that prevent maximum use of test results on the part of the staff.

Using Test Results to Evaluate Curricula

Minimum competency testing has been touted by some writers as a means of assessing the strengths and weaknesses in the instructional and curricular offerings of a school or district. In some programs that have been implemented, the introduction of minimum competency testing has stimulated review of the curriculum in areas in which specific competencies are tested, while in other programs the results have been used as an indicator of the areas in which changes in the curriculum or teaching methods are necessary.

In South Burlington, Vermont, the state mandate that districts assess specific competencies in the areas of mathematics, reading, writing, listening, and speaking led to a review of the grade 1-12 curriculum in those areas prior to implementation of the testing. Administrators and teachers undertook this task in order to determine when mastery of each competency could be expected and hence, when assessments could begin; in doing this they also identified when instruction in each competency begins. Administrators report that staff members share a sense of responsibility for teaching the competencies, since the curriculum review and assessment results have indicated that each grade level, not just the one in which testing of the competency begins, makes a contribution to student mastery of the basic competencies.

In some programs staff members have prepared instructional materials for teaching the competencies in regular classes. Detroit provides one example of a program in which city administrators have prepared a manual that makes suggestions about instruction. Thus, adoption of the competency program there has resulted in additions to the curriculum.

Changes in the curriculum and instructional program have also been initiated because of the results of competency tests. Administrators in Peterborough, New Hampshire, report that staff members have, on their own initiative, reviewed test results and altered teaching methods and course materials when the results revealed major deficiencies in basic skill areas. Administrators and program planners, therefore, may want to con-

sider whether a review of the instructional program and the curriculum, undertaken as part of program development or as a consequence of the test data, is an activity to encourage or initiate.

Using Test Results for State-Level Decision Making

A third way in which test results can be utilized is to assess curriculum and instruction at the district level. In some states, test results are designed for use primarily by state-level officials. In Rhode Island, for example, the implementation of a testing program in basic and life skill areas is designed to provide information to the State Board of Regents on the quality of the educational system as a whole. The Board will use this information in making decisions about the allocation of resources for technical assistance. In Michigan, results of the statewide assessments are used to identify school districts with large numbers of students who are deficient in the basic skills and to allocate resources to these districts so as to correct and prevent these deficiencies. In other states, such as North Carolina, one use of test results is to help in the estimation of the financial assistance districts will receive for remediation. Both the number of students requiring remediation and the severity of their deficiencies are taken into account by state officials in allocating funds. Thus, another way in which test results are used, particularly by state-level personnel in making decisions related to curriculum and instruction, is as a global indicator of the extent to which an educational system has achieved its goals.

In a recent presentation, the Superintendent of Instruction of North Carolina suggested measures to be taken by state and local administrators to encourage the use of test results. He proposed, for example, that reports of test results contain a section devoted to discussing the policy implications of the results. Such a practice can help to ensure that the larger considerations are not lost in the implementation of the program.

Summary

This discussion is intended to illustrate how test results can be used by various groups in making decisions related to instruction and curriculum. Ways to encourage the use of test results in this connection

include informing audiences of how results may be used and the limits of the information yielded by test data. As noted, observers of the rise of minimum competency testing have suggested that administrators and program planners consider issues such as: (1) the relationship of test results to other indicators of the effectiveness of the instructional program, (2) the extent to which the development of competency testing will include or spur curriculum review, and (3) the use of test results in making state-level decisions about providing technical assistance and/or funds.

Options for Organizing Instruction and Remediation

Program planners at the state and local levels have identified a number of arrangements for introducing the competencies to students and for providing remediation. Detroit school officials, for example, have developed a program manual in which they list ways to organize instruction. In addition, other writers have suggested general guidelines for developing competency-based instructional programs, particularly those designed for remediation. This section will discuss options for organizing instruction in the minimum competencies drawn from the work of Detroit administrators and other program planners, noting the guidelines suggested by various writers. In addition, this section will also describe factors that can influence the choice of instructional program.

Creating Special Classes

One way to begin teaching competencies to students or to provide remediation is to create separate classes and instructional materials. Students could attend these classes in order to learn specific competencies, while remaining in the regular program in other areas. Because all students are generally subject to the same competency requirements, program planners who have chosen this option have, in most cases, created special classes for remedial purposes, finding it more feasible to integrate the initial teaching of competencies with the regular instructional programs. Assignment to the special remedial classes is, as in the case of Hillsborough County, an automatic consequence of failing the state or local minimum competency test.

Hillsborough's testing program does provide a cautionary example for administrators and planners who opt to provide remedial instruction through the creation of special classes. In a class action suit brought against both the Hillsborough County School Board, the Superintendent, and various state officials and groups, the plaintiffs claimed that the creation of the compensatory education classes had resulted in a resegregation of the public schools. In a ruling handed down in July 1979, the judge determined that although the classes were populated by a majority of black students, the program allowed the students easy access back into the regular instructional program if they demonstrated mastery of the requisite competencies. Moreover, the purpose of the program was to remedy the educational deficiencies which were a result of previous segregation. An issue, then, that administrators and planners may want to consider, if special remedial classes are created, is how to ensure that students can move easily between remedial and regular instruction.

Establishing Resource Centers

One alternative to special classes is to create centers where students can go for assistance in mastering specific competencies. In Omaha, Nebraska, for example, a student who has missed a specified number of competencies may go to a mathematics laboratory for assistance. Activities in the lab include working with instructional materials geared to those competencies or seeking help from the resource person, who is usually a mathematics teacher on the staff. Administrators responsible for staffing such centers or labs might wish to consider the possibility of employing paraprofessionals or parent volunteers.

Administrators in Detroit suggest utilizing the competency lab to provide more formal instruction to students. For example, lab instructors could teach mini-courses covering one or more competencies for students who were unfamiliar with them or had failed to demonstrate proficiency. Under this arrangement students could remain in their regular classes with the exception of brief periods during which they would attend the lab for instruction.

Tutoring

Tutoring is another way to provide regular competency instruction and/or remediation. Students who have mastered the competencies can tutor students who are just learning the material. It may, of course, be more feasible to introduce all students to the competencies at the same time; in this case, small tutoring groups may not be the most effective strategy to select.

If tutoring is selected as a way of providing remediation, it can occur both inside and outside of regular classes. If the number of students requiring remediation is small, then tutoring might be more practical outside the regular classroom. For example, a nonprofit, nonpartisan organization in New York City, the Public Education Association (PEA), organized a volunteer tutoring program to help New York City high school seniors pass competency tests by June 1979. The competency requirement was the result of a 1976 resolution by the New York Board of Regents, and by February 1979 approximately 15% of the seniors in New York City had not passed the tests in reading and mathematics. The Public Education Association, in conjunction with other interested organizations, recruited and trained adults as tutors. Students were tutored on a one-to-one basis in the high schools during regular school hours when possible. Tutors also utilized other facilities, e.g., community centers and libraries, if needed, and on the average met with students twice a week for one hour. PEA used a variety of media (such as radio, television, leaflets, and newspaper articles) both to recruit tutors and to inform students.

Community centers are the sites used for remedial tutoring in Charlotte-Mecklenburg, North Carolina. In this program, tutorial centers are open after school for interested students. The centers send contact persons to inform students who have failed either the state or local competency tests of the tutoring available at the centers.

With respect to community centers, MCT program planners may want to consider supplementing regular school instruction in the required competencies with tutoring provided by paraprofessionals or volunteers at such centers. Competencies that require practice work or close monitoring in order to achieve mastery might be introduced in the school, but practiced outside of school. Teachers in Peterborough, New Hampshire, developed a booklet on the essential competencies for parents of elementary students. This booklet was designed to explain the particular competencies; it also suggests activities a parent can do at home with the child to facilitate mastery. These activities are intended to supplement the introduction to the competencies a child receives in school.

Individualized Instruction

Another way of providing instruction or remediation that, like tutoring, can occur within the regular program is to have students work independently with self-paced learning materials. These materials can be locally developed for specific instructional or remedial purposes, or state- and district-developed exercises that have been prepared for teaching the competencies may be adapted for remedial use. Both Detroit city administrators and Vermont Department of Education staff members have developed detailed suggestions for teaching the competencies. Local school officials may find their instructional materials useful.

Choosing the Appropriate Arrangements

Factors that will influence the choice of remedial and instructional options include the number of students expected to participate, the size of the instructional staff, the availability of curriculum materials, the physical facilities, the training and interests of staff members, and the availability of paraprofessionals and/or volunteers. As mentioned above, introducing the competencies may be more efficient and cost-effective if done in the context of the regular program of instruction. In cases where mastery of a competency requires close supervision of a student's work or the time spent in class does not permit all the necessary drill, program planners may want to consider supplementing such instruction by using paraprofessionals or volunteers inside or outside of the school. Thus, tutoring would be one way of providing additional instruction, as would providing the students with curriculum materials geared to the competency for independent review.

Real differences emerge when these options, considered as remedial strategies, are compared on the basis of the factors listed above.

Creating Special Classes

Given a fairly large number of students who are approximately similar in ability, compensatory education classes may be the most efficient way to provide remediation. Establishing special classes does entail ensuring that the staff has adequate preparation to provide remedial instruction. This option also entails having sufficient room to accommodate the newly created classes. Demands on the staff could be reduced if paraprofessionals are included as part of the instructional staff.

Establishing Resource Centers

This option makes similar demands upon staff time and the physical facilities. The availability of curriculum materials might help to offset demand for staff time, especially if resource instructors served primarily to refer students to materials rather than to provide actual instruction. Using paraprofessionals or volunteers to staff centers would also reduce demands upon the local staff.

Tutoring

As a way of providing remediation, tutoring may be most effective given relatively small numbers of students needing close supervision. If persons other than teachers or other staff members serve as tutors (e.g., volunteers, parents, peers who have mastered the competency), this arrangement requires a smaller amount of staff time to maintain. The availability of curriculum materials could enhance the effectiveness of the tutors, particularly if they received training in specific remedial techniques.

Individualized Instruction

This arrangement potentially places the least demand upon staff time and facilities. The quality and comprehensiveness of available materials

will, of course, affect the extent to which students will need the assistance and attention of the teaching staff. In addition, staff members may require training in using such materials to teach the competencies. Paraprofessionals or volunteers might also assist students in using self-paced materials.

Summary

The selection of an option is always, of course, the result of trading off factors such as the ones described above. Furthermore, no matter what arrangement is chosen, some writers suggest two more general guidelines: (1) that deficiencies be remediated at the earliest possible instance in the curriculum, and (2) that, if remediation is provided at the secondary level, students be given opportunities to participate in the regular instructional program. The Massachusetts Right-to-Read Committee, for example, asserts that "students must be taught the skills and kinds of knowledge which the tests call for, and remedial instruction must begin as soon as students show they have fallen behind in their progress toward mastery of basic skills" (Slingerland, 1978, p. 12). Speaking to the issue of remediation at the secondary level, Ryan (1979) suggests that remediation be "supportive, not demeaning; that . . . [it] be appropriate to the age level of the student and conducive to the development of self-esteem" (p. 17).

Integrating the Testing Program with Curriculum and Instruction

Apart from determining how test results will be used and by whom, how to encourage their use, and how to structure basic instruction and remediation, program planners and administrators can further ensure the integration of testing with instruction by considering the development of each program component in light of its implications for the instructional program. This section will briefly discuss three specific components--the minimum competencies, the test instruments, and the testing schedule--and how their design can affect instruction and curriculum. The purpose of this discussion is to underscore how a concern with integrating testing with instruction can underlie the entire process of program development.

The Competencies and Instruction

Although the procedures for defining competencies are discussed in another chapter, this chapter will consider this component from the standpoint of the relation between instruction and testing. In some competency testing programs, administrators and program planners have written competencies in order to make them easier for teachers to understand and to teach. In Detroit, for example, city administrators have prepared a program manual in which each competency is carefully defined and ways of teaching the competency to students are described. Similarly in Vermont, State Department of Education staff have developed a handbook for teachers that describes ways of teaching competencies in reading, writing, mathematics, listening, speaking, and reasoning. Competencies written with a view to their comprehension and teachability will ensure that the program components will mesh with the instructional program.

The Test Instruments and Instruction

Testing activities can also be made an integral part of instruction. For example, in some programs evidence of proficiency includes course work or extracurricular involvement. In Omaha, Nebraska, students demonstrate proficiency in problem solving by defining a social problem in a required history course and then proceeding to follow a six-step process to solve it. Steps include proposing and researching a solution. Students first solve such a problem as a class homework assignment, and then choose a different problem for solution in order to demonstrate competency.

In St. Paul, Minnesota, students attending the St. Paul Open School can assemble a portfolio to demonstrate competency in areas such as career education, community involvement, and consumer awareness. Such a portfolio may include letters of testimony from employers and personal accounts of work experiences.

The National Education Association, interested in encouraging the use of indicators other than standardized test results, has cited a variety of options for educators to consider. In their handbook Alternatives to Standardized Testing, Quinto and McKenna (1977) suggest contracts, conferences, and teacher-made tests as ways of assessing proficiency. While the authors address the more general issue of how to assess student progress, their discussion is relevant to the issue of how the minimum competencies may be assessed. Their suggestions for alternative means of assessment may provide a way to better integrate testing and instruction.

The Testing Schedule and Instruction

To give yet another example of how a program component can be viewed in terms of the relationship between testing and instruction, consider the issue of how to determine the testing schedule. In Vermont, for example, the state specifies the competencies to be assessed but does not specify the testing schedule. Rather, the state stipulates that beginning with the class of 1981, students must master competencies in particular areas in order to graduate. To determine when to begin assessing students on the basic competencies, administrators in South Burlington, Vermont, along with a group of teachers, conducted a curriculum review. The purpose of the review was to find out when instruction in each competency began, and to estimate when a student could be expected to have mastered each competency. The point at which mastery is expected is the point at which the student is first assessed on the competency. In this program, testing activities were keyed to the instruction.

In addition to considering the option of relating the testing schedule to instruction, administrators may also promote the integration of testing and instruction through carefully weighing the potential impact of the testing schedule upon curriculum and instruction. One such consequence, if the numbers of students requiring remediation is high, might be a need for teachers with special training in teaching the competencies at an appropriate level. For example, since introducing minimum competency testing into the high schools, school administrators in Gary, Indiana, have hired teachers who are trained in providing remediation in the basic skills to high school students. These administrators discovered that many secondary teachers either were not trained to teach basic skills in high school or were not interested in teaching remedial classes. Instituting remedial classes at the high school level meant hiring teachers specifically to teach remedial courses in basic skills such as reading. Thus, the selection of a testing schedule may result in special demands being made upon the talents and interests of the staff.

Summary

The above examples are intended to illustrate the point that a concern for strengthening the relationship between testing and instruction need not be limited to considering how to promote the effective use of test results and possible remedial strategies. This concern is an appropriate one for all stages of program development.

References

Detroit Public Schools. Detroit High School Proficiency Program: Program manual. Detroit, Michigan: Author, 1979.

Goslin, D. A., Epstein, R., & Hallock, B. A. The use of standardized tests in elementary schools (Second Technical Report). New York: Russell Sage Foundation, 1965.

Means, H. J. Reading and minimal competency testing. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Phillips, A. C. State and federal roles in testing: As viewed by the state superintendent of North Carolina. In R. M. Bossone (Ed.), Proceedings of the Second National Conference on Testing. New York: Center for Advanced Study in Education, 1978.

Quinto, F., & McKenna, B. Alternatives to standardized testing. Washington, D.C.: National Education Association, 1977.

Ryan, C. The testing maze. Chicago: National PTA, 1979.

Shoemaker, J. S. Minimum Competency Testing: Implications for Instruction. (Unpublished paper) Washington, D.C.: National Institute of Education, 1979.

Slingerland, J. The minimum competency movement. Chairman's Report, Massachusetts Advisory Council for the Right-to-Read Effort. Boston, 1978.

CHAPTER 6

PROGRAM MANAGEMENT

William Phillip Gorth and Peter E. Schriber*

Introduction

This chapter will present a set of preliminary procedures for preparing a management plan for a minimum competency testing program. In the preparation of such a plan, the specification of personnel needs and the determination of costs will play significant roles. Since budgetary constraints affect every component of a program, this discussion will touch upon the nature of the costs which an MCT program is likely to entail. It should be stressed, however, that neither specific costs nor estimates will be offered to the reader.

In addition, those responsible for the planning and management of an MCT program will find it necessary to locate and identify personnel to perform the many tasks which the program may require. Consequently, guidelines and strategies for meeting personnel needs will also be discussed here.

This chapter essentially provides a repertory of procedures and strategies from which educators responsible for program management can draw at will. This presentation will not exhaust all possible alternatives and will not prescribe specific techniques or modes of organization. It is a possibility that none of the procedures under discussion will be apposite to a particular program. It is hoped, however, that even in such a case, this discussion will be useful in that it may stimulate educators to look at their management needs in a fresh light as a result of the considerations introduced here.

The topics brought forward in this chapter have been selected because they are the issues which seem to be of the greatest interest or concern to those responsible for the design of management plans for competency programs. In the course of the discussion, examples will be drawn from

---

* With organizational assistance from Dolores R. Harris.

various programs. This practice, however, is in no way an endorsement of a particular procedure; these examples have been chosen only to illustrate a point more clearly and to suggest the wide range of solutions possible for each of the problems or topics under examination.

As a step preliminary to planning for an MCT program, it has been found useful to establish a center of control or focus of responsibility for the activities which are to be undertaken. In all 52 of the programs of the study--whether initiated at the state level or the local district level, whether initiated by legislative mandate, by the action of a state board, or at the direction of a local agency--the control and administration of the program had been delegated to a single agency or individual that assumed all responsibility for planning, coordinating, and managing all the activities which the program called for. In the field, a variety of arrangements for this purpose were encountered. Throughout this chapter on planning it will be assumed that the center of control and responsibility for an MCT program has been established, and, for the sake of convenience, it will be assumed that this center of control resides in the person of a program director. However large or small the program, the duties and functions of such a program director remain essentially the same from program to program.Since this role is such an important one, it may be worthwhile to consider the functions of the director and the qualifications which might equip a candidate to occupy this position and carry out its duties successfully.

In a minimum competency testing program the director occupies a position which is intermediate between the initiating or policy-making bodies, and the constituencies that will be affected either directly or indirectly by the program. Important qualifications for the director, therefore, may be the ability to understand the diverse viewpoints and concerns of these groups, conjoined with the ability to find common or unifying themes in this diversity which will facilitate the task of implementing the program. To increase the likelihood of accomplishing the program goals, the director might best be drawn from a pool of candidates familiar with a given educational system and with the community it serves. Experience in educational planning and administration and demonstrated ability to organize and direct groups are also extremely desirable attributes in a prospective program director.

As additional sources of information on the duties and qualifications of a program director, the reader may wish to consult the Competency Handbook of the Ohio Department of Education and the California Technical Assistance Guide for Proficiency Assessment, both of which appear in the reference list at the end of this chapter.

In order to assess the planning and management needs of a minimum competency testing program it may be helpful, as a first step, to prepare a full account of the stated purposes and goals of the program and a list of the prescribed activities through which these goals are to be realized. This procedure can clarify the nature and extent of the task, since it will delineate the essential and irreducible structure of the program. This essential structure or form will, of course, vary from program to program. In some instances, as in certain statewide programs, for example, the policy-making body has not only initiated the program, but has also specified its components in detail. Such programs may present the planner with a set of competencies, an established testing schedule, predetermined target groups, approved testing instruments, prescribed standards, and explicit directions for generating reports of the test results and for the uses of these test results--both in making decisions about students and in supplying information to the public.

At the other extreme, some state and local programs have been formulated in the broadest terms possible, leaving decisions on these and other issues to the discretion of the individual agency, and, in effect, to the program director or planner. In either case, however, this first procedure will establish all and only the essential elements of the MCT program.

It may be useful at this point to categorize these elements as belonging to one of three program components: (1) instruction, (2) testing, and (3) remediation. By definition, all MCT programs will have a testing component. The inclusion of one or the other, or both, of the remaining two components appears to be an optional feature. It may be wise to point out here that the adoption of this mode of categorization does not mean that all of the components in a program are of equal importance. For example, one program may key its testing to the curriculum, so that the curriculum components will define the domain of the testing components; this has been the course followed in the MCT program in South Burlington, Vermont. In yet another program, the reverse may be true: the testing component will establish the desired educational results, requiring the adjustment or redesign of the curriculum component. The program in Peterborough, New Hampshire exemplifies the second configuration.

After the essential structure of the program has been outlined, its various elements categorized under the appropriate components, and the hierarchical order of the components determined, it may then be possible to specify the tasks necessary to implement each component. To characterize the nature of each task identified, it is often helpful to ask a set of questions which will determine the procedures and resources necessary to accomplish that task. For the purpose of discussion, it will be assumed that the task is a unitary one, which cannot be broken down into subtasks. Some appropriate questions might be grouped as follows:

C295

SEQUENTIAL ORDER:

-- What tasks, if any, must precede this one?
-- What tasks must follow?

METHOD:

-- What methods can be employed to accomplish this task?
-- What methods are available for use in this program?
-- What method is the most feasible for this program?

RESOURCES:

Personnel

-- What personnel does this task require?
-- What personnel are available?

Expertise

-- What kind of expertise does this task require?
-- What expertise is available?

Time

-- How much time will the task require?
-- How much time is available?

Funds

-- What expenditures are necessary for the task?
-- What funds are available?

These questions point up the fact that in planning and managing a program, the issue of program needs versus the availability of resources needs to be considered at every step.

Since specific tasks, such as identifying the competencies, test development, standard setting, and dissemination, are dealt with in other chapters in this document devoted to these topics, the remainder of this chapter will concern itself with a discussion of two subjects: personnel resources and the ways in which a program director might develop and employ these resources to their maximum effect in order to achieve the stated goals of the program; and the costs which an MCT program may involve.

Personnel

A program director may wish to call on both internal and external sources to satisfy the personnel needs of a given MCT program. Internal sources include the teaching staff, administrative and technical staff, and clerical staff employed by an educational system. These staff members will be the most likely source of personnel for tasks which, for their accomplishment, require specific knowledge of content areas, methods of testing and evaluation, and curriculum design. In some programs, the local district may also be in a position to draw upon the expertise of state-level specialists to assist them in these matters. Internal staff often play an important role in test development in programs which engage in this activity. Also, members of the teaching staff usually administer tests, and frequently score them. Remediation, reporting results, and dissemination are other activities in which internal staff may participate, conditional upon the design of the MCT program.

External resources for personnel may include outside educational contractors, consultants, or specialists called in to assist with one or more components of the MCT program. Their use is often dependent in large part not so much upon need as upon the availability of funds for this purpose.

A very important source of external personnel is, of course, the community which an educational system serves. It appears that the most successful programs of the study, in many cases, were those which engaged a broad representation of community members in the tasks of program development. Active involvement seems to generate support and enthusiasm for a program which can act as a powerful catalyst.

The formation of a committee is the most usual method by which community members are drawn into active participation in an MCT program. A review of the programs of the study will reveal the wide variety of activities and tasks which such committees have undertaken in the design and implementation of MCT programs at both the state and local district levels. And it has been observed that "communities are more prone to accept changes in their school systems if they are not only informed but also involved in the process" (California, SDE, 1977, pp. iii-6).

There are at least three kinds of committees which can be employed in MCT programs:

(1) ADVISORY COMMITTEE -- represents wide range of interests and reviews general program policy.

(2) STEERING COMMITTEE -- deals with detailed aspects of program policy, may prepare draft versions of policy statements, and may have membership which is a subset of the Advisory Committee.

(3) WORKING COMMITTEE -- one or more Working Committees may be established to accomplish specific tasks necessary to implement the MCT program and may have membership which overlaps partially or not at all with the Advisory Committee (Ohio, SDE, 1978, pp. 2-4).

A review of the state materials prepared to assist planners with MCT program development shows widespread agreement on the considerations which are especially relevant to the formation of such committees.

Committee Composition

It is recommended that the composition of the committee be carefully planned, appropriate to the tasks it will be assigned, and representative of the community affected by its work, whether that community is defined by the geographical boundaries of the school district or of the state. The committee, if it represents a cross section of the community, can make it possible to gather information about all constituencies as to what they want, approve of, understand, and will support. The extent to which these different constituencies are involved in the MCT program may determine the extent to which the results of the program will be supported by the community. It is important to realize that special interest groups within the school may be as important as those in the community; therefore, the members of such groups will also be desirable as committee members.

Committee Selection

It is recommended that a selection strategy be explictly determined by the local or state board or superintendent and implemented by the program director. One strategy is to appoint individuals who have been active in school affairs. However, if all the members are selected in this way, the committee may not accurately reflect the community. A second strategy is to set guidelines for selection in order to achieve a

balanced membership. A third strategy is to solicit the participation of community members by open invitation, which allows for greater community involvement. It is possible also that a combination of these strategies can help to obtain members from all the interest groups crucial to program success. At the beginning of the selection process it may be wise to emphasize to prospective members that they will be expected to serve actively.

Committee Functions

The functions of the committees versus those of the program director may need to be clearly differentiated. Although each committee may be generally considered as advisory in nature, a committee can assume a decision-making role as a primary voice to the community or as the technical experts in a particular subject. Therefore, the committee may be useful as a forum for sounding out ideas or for defining and selecting alternative approaches at every stage of program development.

Committee Size

The nature of the task which a committee is to perform will very often determine its size. If the committee is an Advisory Committee designed to represent community interests adequately, it may very well contain 25-50 members (e.g., Massachusetts Statewide Advisory Committee). A Steering Committee, on the other hand, may require only 5-10 members to handle material development effectively (cf. the Detroit Public Schools program). Working Committees usually require 5-12 members to represent the various professional opinions adequately.

* * * * * * * * * *

For further information on this subject, the reader may wish to consult the materials prepared by the California, Illinois, and Ohio Departments of Education. These handbooks present useful information, organizational charts, and strategies designed to assist program planners in meeting the personnel needs of their programs.

It is also possible to achieve community participation by other means. In some programs public meetings have been convened for the purpose of permitting members of the community to express their views about the minimum competency testing program. If possible, several such meetings could be held in different locations in order to reach as many people as possible. Meetings may also be scheduled for such groups as business and professional organizations, trade unions, associations of parents and advocates of students with special needs, and ethnic and cultural organizations. It is advisable to prepare carefully for such presentations, since they will usually serve a dual function; not only do they permit educators to collect information about the concerns and needs of community members, but such meetings also provide the educators with an opportunity to inform the public about the goals and purposes of the MCT program.

The survey is another useful method for reaching the public. It may be a comprehensive survey, such as that employed in the Detroit program, in which completion forms printed in the local newspapers solicited the opinions of all those wished to respond. On the other hand, a survey can be employed to focus on a particular segment of the community. In the Maine program, the Benchmark Survey was confined to a representative sample of high school teachers, and sought their views on the performance levels which could be reasonably expected of Maine eleventh-graders.

The public meetings and surveys described above were connected with various aspects of program design and development. Another task to which members of the community might contribute is that of remediation. In New York City, volunteers were recruited and trained as tutors for deficient students. Such a measure has the added advantage of supplying students with individual remedial instruction at a relatively low cost. Remediation offers opportunities for involving certain other constituencies with an interest in the program. Parents, of course, have an obvious interest in a child's success, and many programs require parental participation in the design of remedial programs for a student who has failed the minimum competency test. Parents who are involved in this fashion may well be more receptive to suggestions as to how they may help their children to achieve mastery in the required competencies.

The Detroit Program Manual suggests peer tutoring as one way of meeting the remediation needs of an MCT program. Students with demonstrated competency may be able to assist their contemporaries to acquire the skills needed for mastery, and deficient students may respond more positively to instruction from a fellow student. Program directors will know best which strategies are appropriate for use in their own programs.

Costs

The costs of a program will be dependent upon the components of the plan. For example, a program may be designed as part of a larger competency-based educational program, in which new curricula are developed, or it may entail the use of a single, commercially available test. Obviously the costs for each program will differ greatly. Therefore, it may be most useful simply to characterize the various kinds of costs common to most programs.

Airasian, Madaus, Pedulla, and Newton (1979) discuss costs associated with MCT programs under four different categories: program development, test administration, consequences of the program, and intangibles relating to acceptance of the program. The following discussion has adopted the first three of these categories in order to present the material systematically.

Program Development

This category covers start-up costs. They occur only once in a program; however, if the program is constantly refined, these costs may have their counterparts in the maintenance costs of the program.

Planning. These are largely personnel costs and at the local school district level may be absorbed in regular salary time by the reallocation of staff efforts. However, the more complex the program, the higher the cost for staff because a complex program will require more staff time for planning.

Identification of competencies and development of competency statements. In most programs this process will involve input from educators and community members. Time of the program staff is necessary to coordinate the activities of the many people serving on the advisory committees and working committees which are usually involved in the development of competencies. Because the competencies are the basis for later development, their identification may require a relatively large amount of staff time.

Development of curricula or matching competencies to existing curricula. This includes the alignment of instruction with the MCT program. If curricular development or modification is planned, significant costs may be necessary to fund staff time for the development of materials, duplication, and secretarial support.

Program dissemination. Supplying information to the community and the staff of the school may be one of the requirements of the program. Staff time will be necessary to write the notices and reports. Printing and distribution is directly proportional to the number of persons contacted. Nonprint media may be much more expensive to develop but have a lower distribution cost, if radio and television stations will contribute the time.

In-service education. Staff members may need training in developing/selecting competencies, interpreting and use test results, and in planning instruction to align their teaching with the competencies. Costs may be separated into preparation of materials (professional staff, secretarial support and printing) and costs for the presentation (presenter and participant time).

Test selection/development. More staff time is necessary to develop than to select a test. A commercial test, however, involves the cost of buying copies of the test for each administration. Either development or selection will require staff time to consider the content appropriate to the test, and to review the test with committees. If the test is developed, added staff time will be necessary for writing items, editing items, pilot testing items, revising items, and producing the final copy of the test for duplication. Supplies, the duplication of materials for review and pilot testing, support for the analysis of pilot testing, and secretarial support for the preparation of drafts and final copy will also be necessary.

Testing

After the test has been selected or developed, a number of costs will be repeated at each administration. These costs are stable from year to year, except for increases due to inflation, and, therefore, predictable.

Test administration and scoring. Space and time allocation, test administrators, test printing or purchasing, test security, test distribution and collection, and test scoring all require an estimation of costs.

Reporting of test results. Preparing and writing the reports of test results, whether computer-based or narrative, for the student, the parent, the media, and for instructional staff will result in expenses for staff time, secretarial time, printing, and distribution. Computer programming, computer time, and consultant time will add to the expense if the agency feels it needs these resources.

Provisions for special students. Students with special needs and limited English-speaking students may entail additional costs, if the program decides to offer alternative assessment strategies for these students.

Consequences of the Program

Instructional implications of the testing results. The available resources of money, teacher time, and instructional materials will determine the number of students served and the nature of a remediation or alternative instructional program.

Litigation. Since an MCT program focuses on student performance, some lawsuits have been filed with respect to the legal grounds of such a program and its policies. Contingency planning for the costs of staff time and legal services in this connection may be necessary.

Dissemination. Test results are important to the public in their function both as parents and as taxpayers. Expenditures which may be involved in the dissemination efforts of a program are discussed in detail in the next chapter.

* * * * * * * * * *

In addition to the discussion in this chapter, a monograph published by the U.S. Office of Education, titled The Resource Approach to the Analysis of Educational Project Cost, presents a model which is based on

the resources necessary to operate a program and which may be used to compare different configurations of a project in different locations. It may be useful to make a preliminary estimate of costs based on the information provided by existing programs at district and state levels. Other articles which provide general information about costs are Anderson (1977) and Miller (1978).

C295

References

Airasian, P., Madaus, G., Pedulla, J., & Newton, K. B. Costs in minimal competency testing programs. In P. Airasian, G. Madaus, & J. Pedulla (Eds.), Minimal competency testing. Englewood Cliffs, New Jersey: Educational Technology Publications, 1979.

Anderson, B. The costs of legislated minimal competency requirements. St. Louis, Missouri: CEMREL, 1977.

Brickell, H. M. Seven key notes on minimal competency testing. In B. S. Miller (Ed.), Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

California, State Department of Education. Proficiency assessment in California: A status report. Sacramento, California: Author, 1979.

Communication Technology Corporation. Project director's manual. Moreton, New Jersey: Author, 1976.

Dalton, S. Commercial versus school-district-made tests to measure minimal proficiencies for high school graduation. Riverside, California: Riverside Unified School District, 1978.

Gourley, R. N. Learning to manage a minimum competency testing program. Paper presented at the meeting of the American Association of Colleges for Teacher Education, March 1979.

Illinois, State Office of Education. Establishing educational priorities through the Illinois Problems Index: User's manual. Springfield, Illinois: Author, 1977.

Miller, B. S. (Ed.) Minimum competency testing: A report of four regional conferences. St. Louis, Missouri: CEMREL, 1978.

C295

New York, State Education Department. Regents examination manual. Albany, New York: Author, 1976.

Ohio, State Department of Education. Competency handbook. Columbus, Ohio: Author, 1978.

The resource approach to analysis of educational project cost (Evaluation in Education Monograph No. 3). Washington, D.C.: U.S. Government Printing Office, 1978.

Shepard, L. A. A method for evaluating assessment. Paper presented at the sixth annual Conference on Large-Scale Assessment, Boulder, Colorado, June 1976.

C295

CHAPTER 7

DISSEMINATION

Peter E. Schriber and William Phillip Gorth

Introduction

Purpose

This chapter discusses issues and techniques relevant to preparing a dissemination plan for a minimum competency testing program. The issues and techniques are those identified in the survey of 31 state and 20 local MCT programs; the discussion is based upon interviews with program planners and administrators as well as an analysis of program materials. In addition, the writings of other professionals in educations were used to highlight key points.

The considerations and suggestions presented below are neither exhaustive with respect to the general topic of dissemination nor are they prescriptive in nature. The discussion is directed towards program planners and presents examples of considerations and practices they may wish to consider in developing a dissemination strategy or an MCT program. Existing programs and program materials are cited to illustrate specific points.

The chapter is organized in the following way. The basic elements involved in the planning process for dissemination are presented and discussed first, with examples of ways in which such a plan may be documented concluding the chapter. The significant outcomes of the planning process are the identification and selection of appropriate media by which dissemination of information about an MCT program is to take place. While the discussions of these outcomes appears late in this chapter, it essentially serves as the justification for the earlier discussion.

The Planning Process

Program planners suggested that dissemination activities be carefully planned and executed in order to maximize effectiveness. Since a major purpose for dissemination is to promote awareness and gain acceptance and support for an MCT program, a poor dissemination effort may result in strained community relations, misunderstanding by special interest groups both in the community and on the school staff, and loss of community trust in the schools. The following subsections discuss issues raised by program planners, as well as those found in program materials.

Identifying the Purposes for Dissemination

It is assumed that the MCT program is necessary, endorsed by the schools in the district (or, if statewide, by the districts in the state), and so designed as to achieve its objectives. In general, it has been found that, if the program initiators and implementers are not behind an MCT program, the dissemination effort is likely to be of little use. For instance, the Charlotte-Mecklenburg, North Carolina, district carefully planned approaches for community awareness and involvement through public media and other community outreach efforts from the very inception of its MCT program. This dissemination effort was and continues to be an important activity of the managerial staff of the program. In general, dissemination activities are an integral part of an MCT program and, as such, require as much thorough consideration and planning as other program components.

Discussions of the general purposes and principles of dissemination may be found in various materials. The California Department of Education in its Technical Assistance Guide stresses the importance of using dissemination to promote community involvement, while the National School Public Relations Association publishes a booklet on this topic. Goals of dissemination identified in this publication, as well as in the California material, include to inform, to gain acceptance or compliance, to obtain support, cooperation or participation, or to encourage the use of results.

Since the first three purposes typically require increasing degrees of involvement on the part of the target audience, each succeeding one may require more effort to accomplish.

Identifying Types of Information to be Communicated

California administrators suggest that a critical component of a dissemination plan is usually a specification of the types of information to be disseminated, and as a result, that program developers consider the kinds of information various audiencies may be interested in. Consequently, it may be advisable to start by compiling a complete list and by organizing this information list in such a way as to achieve an overall view of the dissemination plan. This will help to identify and remedy any major gaps that may be apparent in the plan.

It may also be helpful to prepare a detailed description of the MCT program, complete with the rationale for each component and procedure, for eventual communication to different school and community audiences (Hubbell & Stech).

Identifying aspects of the program. Among the issues typically considered by program planners in determining what to disseminate are the amount and the nature of the information to be provided. Current MCT programs across the country generally recognize that the purpose of a dissemination effort is to present a coherent view of a well-designed and well-conceived program with clearly expressed goals which do not discriminate against any group. The major aspects of MCT programs are identified below in the form of a checklist which may be useful to include in a dissemination plan. The specific details of the plan for a local or state-level program may be completed as the planner sees fit.

(1) Program name
(2) Policy history
(3) Program goals/purposes
(4) Competencies
(5) Standards of performance
(6) Target groups and testing schedules
(7) Test instruments
(8) Test administration
(9) Use of test results

Determining the types of test results. Test results can be generated and reported in many formats and in various forms of descriptive statistics. Understanding these results and the different modes of presentation for these results is often a problem for the disseminators as well as the intended audiences. A thorough discussion of the types of results that can be prepared for particular tests and audiences is beyond the scope of

this document. However, being aware of the importance of test results and of their impact on various audiences will facilitate the planning for dissemination. Test results are the data most easily and most often misunderstood in any program which involves testing (Hubbell & Stech). One important reason for this is that numbers, scores, and statistics may be reported either without sufficient explanation or without sufficient knowledge of the level of understanding that each audience will need in order to assimilate the information in the manner prescribed in the dissemination plan.

A discussion of alternative methods of presenting test results to various audiences and for the use of test results in instructional diagnosis and planning is presented in another chapter of this document.

Identifying Key Target Audiences

The MCT programs represented in the study disseminate a wide variety of information to a wide variety of target audiences. In general, it has been found that a well-planned strategy will identify these audiences and select the information and the dissemination method appropriate to each. To assist in this task, some MCT programs, such as California and Florida, have asked the following questions:

-- What are the audiences and who are their members?

-- What are the critical concerns of each audience?

-- What is the perspective of each in understanding or dealing with the MCT program?

-- What information must be presented to each audience and for what purposes?

-- How will critical concerns be faced?

In planning a strategy to answer these questions, there are specific issues and guidelines which help to add focus. The discussion which follows is based on discussions with program personnel and on treatments of the issues by Hubbell and Stech in a publication of the Colorado Department of Education. Critical issues and decisions will be highlighted and potential problems in the dissemination effort will be identified.

As an introduction to the discussion, it may be useful to consider the various audiences for information from which support may be needed. There are many audiences, and it may be important to identify as many groups and key individuals as possible. For instance, special interest groups may have particular and potentially troublesome concerns about the MCT program. It could be crucial to identify each of these groups and to anticipate its concerns, since the support of special interest groups may make a program. Similarly, the absence or withdrawal of this support can break it.

Some audiences may become more involved or concerned over time. An MCT program can run for a considerable length of time, and may even become a permanent program. Just as the program may be modified and revised over time, so audiences will change in composition and particular interests. New special interest groups may emerge. For example, local businessmen may come to depend on test results in hiring high school graduates. Parents whose children are preschoolers at the outset of a program will take a greater interest when their children participate in the program.

There are certain subgroups which may require special attention in the dissemination plan. It should perhaps be kept in mind that such an audience need not be large to be essential to the success of the program. For example, the town council and local labor unions can be small in size but extremely influential. Neighborhoods with predominantly non-English-speaking residents may need special consideration. A language or socio-cultural barrier may mean that a special effort is necessary to keep all the people in a community fully informed and to keep communication problems at a minimum.

In identifying all pertinent groups, it is important to recognize that there are groups within the school system itself which are also potential target audiences: teachers, students, and school administrators.

Since an MCT program is very likely to be labeled an "assessment" program, it is also likely that many audiences will have a personal and emotional interest in such a program. These groups may feel that their student members will be stigmatized or discriminated against as a result of their performance in the program.

It is useful to remember that the MCT program is essentially for everyone's benefit. But, since it is a testing program, it will identify students as deficient. And, as any program, it requires tax money for its support. These two facts alone may generate negative feelings which a cursory or half-hearted dissemination effort will do little to allay.

Identifying Audience Concerns and Goals of Dissemination

The following list presents a sampling of categories of target audiences with a brief characterization of the typical concerns and the goals planners might set in developing a dissemination plan. Programs in which some or all of the audiences listed below have been identified and addressed through various media include, for example, Michigan, California, Florida, and North Carolina. Program materials from Florida and North Carolina include pamphlets to students dealing with some of the concerns listed below. Michigan educators focussed on a number of audiences, including district administrators, in developing dissemination materials, while California, in its Technical Assistance Guide, describes how an administrator might address the concerns of the community in presenting assessment results. The discussion below, then, is based upon interviews with planners as well as an analysis of program materials. For other discussions of the same topic, the reader is referred to Hubbell and Stech (n.d.) and NSPRA (1976).

(1) IN-SCHOOL AUDIENCES. The four major categories are: students, teachers, administrators, and boards of education.

(a) Students

Concerns: Consequences of poor test performance are usually the chief sources for concern. Questions most frequently asked are:

-- What happens to me if I fail?

-- Do I get behind in other courses if I am assigned to remediation?

-- Is there a stigma attached to being in a remedial group?

Dissemination goals: Gaining student acceptance and allaying their fears are foremost. One approach which may be employed is to help students understand that remediation will make them more employable and better prepared to face life after graduation. Planners may also want to give considerable thought to the means by which passing or failing scores are reported.

(b) Teachers

Concerns: Teachers may feel that the program will add to their workloads. Some may also feel that differences in test scores among classes will lead to evaluations of teacher performance. Other concerns are the effect of the program on the students and potential curricular changes. Questions may include:

-- What additional duties will be expected of me?

-- Will the administration rate me as a teacher on the basis of my students' test scores?

-- Will the program be beneficial to students?

-- Will the curriculum be changed? Should the curriculum be changed?

Dissemination goals: If the program is a local one, many of these concerns may be addressed by encouraging teachers to take an active part in the formation of the program. It is common knowledge that a program has a better chance for success if the participants have planned and developed the program themselves. Thus, beyond mere acceptance of the program, teachers may be more supportive of the program if they are active participants.

(c) Administrators

Concerns: These will vary from administrator to administrator. Primary concerns may include a loss of operating funds due to the fact that program needs have received priority, extra work involved in organizing staff for program implementation, and impact on the daily school routine of program components which must be scheduled. The administrator may also be concerned about comparisons of schools based on test performance, about the administrator's role in directing program components, and about increased community concern translated into more frequent requests for information directed to the administrator. Questions include:

-- Will I lose funds for my school because of money put into the program?

-- Will I have extra work to do in terms of planning for testing (or for curricular or test development)?

-- Will my staff have extra duties to perform?

Dissemination goals: Gaining acceptance and obtaining cooperation and support are considered to be key goals in terms of having every school in a district participate equally. District-level and building-level administrators may require thorough briefing on their roles and the roles of their instructional staff in MCT program development and implementation. As with teachers, active participation may foster cooperation and support. The California State Department of Education, in its handbook for local school participation in MCT programs, delineates roles for administrators within the MCT programs such as program monitoring, involvement in standard setting, and establishing remediation courses and alternatives.

(d) Boards of education

Concerns: The greatest concern is community impact. Questions include:

-- Will the community provide positive support?

-- Will the program better prepare students for life after graduation?

-- What special interest groups may respond negatively?

Dissemination goals: The board of education may need to be involved from the inception of a local district program. The board is very likely to expect information about the dissemination efforts planned for the other target audiences.

(2) COMMUNITY AUDIENCES. These audiences are parents, residents without school-age children, business groups, and special interest groups.

(a) Parents

Concerns: Their main concern is generally for the effect the program will have upon their children. This concern is often manifested as fear or anxiety that the MCT program will single out for failure the students with learning problems and other disabilities. The way in which the issue of parental concern is handled can play a significant role in determining the success of a program. Questions include:

-- What are the criteria for passing or failing?

-- Will my child get special attention if he/she doesn't pass the test?

-- Will a child who fails be singled out and stigmatized?

-- Will the program focus on weaknesses in school academic programs?

Dissemination goals: Program directors agree that parental involvement in and support of the program is essential for its success. Parents need to know why the program has been initiated, what it will test, and why. One device commonly used by many local districts is a parent council to review program content so that parental understanding of the content of a program is maximized. The Michigan Educational Assessment Program adopted another approach and produced several question-and-answer newsletters and brochures for the general community. One, entitled A Pamphlet for Parents, is directed solely at parents and describes the program, lists sample objectives, and provides information about the standard of performance. Careful attention to providing information on how students with failing test scores are treated and remediated is important in every current program.

(b) Residents without children in school

Concerns: These people may be members of several different target audiences. They can be childless couples, parents with preschoolers, parents with grown children, and elderly or fixed-income people. Their concerns may range from the amount of tax money necessary to the impact of the MCT program on the community. Questions include:

-- Will taxes go up?

-- What good is more testing?

-- How will students be better educated because of the program?

-- Will the program reduce the number of graduates who are functionally illiterate or unable to perform simple arithmetic calculations?

Dissemination goals: The general purpose is to gain community acceptance and support. Fears about increased taxes may need to be allayed. One possible approach is to clearly describe the benefits of the program to the community. The American Friends Service has put out a guide for the general community, entitled A Citizen's Introduction to Minimum Competency Programs for Students, which describes succinctly and clearly what citizens look for in developing and evaluating MCT programs.

(c) Employers and business organizations

Concerns: A chief concern is whether the program will prepare graduates better for entrance-level occupations which require only a high school diploma. The most frequent question is:

-- Will the students passing the test make better employees?

Dissemination goals: Acceptance and support of the program may be facilitated by showing a connection between school preparation and success on the job.

(d) Special interest groups

Concerns: These groups may include trade unions, socio-cultural neighborhoods, ethnic-identity groups, or equal-rights groups. A major concern of such groups is the possibility that the MCT program may discriminate against members of the group. If a particular group has a disproportionately large proportion of students who have failed the MCT tests, then to achieve success, the MCT program will need to engage the support of the parents of the deficient students in order to help such students to participate and succeed in the appropriate remedial programs. Failure to meet this issue might lead to charges of discrimination, because the special interest group may feel that the MCT program is designed only to label its student members as deficient; the group may need special attention to see that its members understand the function of the remedial component of MCT as well. Questions include:

-- Will failure on the test reinforce a student's negative feelings?

-- Will the program stigmatize minority groups?

-- Why is the program good if students are rated by test scores?

Dissemination goals: Gaining acceptance is a first-level goal. Cooperation and support would be greater if any of the members of the special interest groups have children who will be involved in the MCT program. One apparently effective approach to this problem is illustrated by the community outreach program implemented in the MCT program of Charlotte-Mecklenburg, North Carolina. As part of its remediation effort, the district has organized after-school community tutoring centers in disadvantaged neighborhoods and has initiated a door-to-door outreach effort to inform parents of students who fail the MCT test about the remediation program and its value to their children.

(3) OTHER AUDIENCES. Important audiences that may go beyond the boundaries of a community are the news media and education associations.

(a) News media

Concerns: The news media may be the special group which requires the most careful dissemination effort of all. The members of this group may see the program as an additional source for news, and test scores as interesting reading; unless there is good rapport between the district (or the state) and the media, anything which is controversial may be emphasized at the expense of the goals and successes of the program. Questions from the media may include:

-- What is a good test score?

-- What is the rationale behind the program?

-- Are the program goals realistic?

-- What is the response of various interested groups?

-- What are the consequences to students?

-- Is the program well conceived and implemented?

Dissemination goals: Many program contacts agreed that good press support can be essential to make sure that incorrect or distorted information is kept to a minimum, and to gain public support of the program. For a description of ways administrators might present information to media representatives as well as a sample news release, see California (1977). Many programs plan several news conferences and even hold public question-and-answer sessions on television (as, for example, the Charlotte-Mecklenburg, North Carolina district). The Florida State Department of Education invited 37 representatives of the news media to take one of its minimum competency tests and then to write stories about their impressions of the test (Fisher, 1978).

(b) Education associations

Concerns: Education associations, including unions, may view the MCT program as a threat to the teacher--in terms of the extra, generally uncompensated work the program may require, and in terms of the potential for teacher evaluation which may be based unfairly on student test results.

Such groups may ask whether the program really has been designed to aid students or to provide a cosmetic and superficial means of satisfying the community concern for accountability in education. Questions include:

-- Will the program really help students?

-- Will the teacher face an unfair burden of extra work?

-- Will any part of the program evaluate a teacher's performance?

Dissemination goals: If the MCT program is statewide, state-level education organizations can be key target audiences, and the goal of dissemination might be to gain program acceptance. A local program may have to win over the local teacher associations and unions. If teacher support has been fostered at the grass-roots level, then organizational support or acceptance may be easier to obtain.

Identifying Resources for Supporting Dissemination

Dissemination is a large task which may require a substantial commitment of time and money. Two critical tasks for a dissemination plan are to determine the message and to get it across to the right audience. The chief resources of a dissemination effort are its personnel and the means available for reaching the various audiences.

Personnel. In planning for dissemination, involving key people from the earliest planning stages of the MCT program (Hubbell & Stech; NSPRA, 1976) can be important. These persons are most likely to be extremely familiar with the MCT program and in close touch with the political leadership of the community. It might be useful if the release of information is monitored by the state or local district administrative leadership. Public relations always play a large part in the operation of any educational agency. The personnel in charge of dissemination can enhance the chances for the success of a program if they are experienced and of sufficient stature to command respect of any groups or constituents they may have to address.

Often dissemination is a team effort. At the state level this team may consist of State Education Department public relations staff and MCT program staff. At the local district level, such a team might include the superintendent or a designate, the MCT program director, and guidance staff--all persons experienced in community interaction within the school district.

Materials and funding. Although it is impossible to develop a formula for the funding of a dissemination effort, two useful activities are allowing for the allocation of funds and setting up a budget for this purpose (NSPRA, 1976). In this connection, a major factor to consider may be the nature and the amount of effort necessary to ensure adequate acceptance and/or support of the program from the target audiences.

To offset the large costs of dissemination and to handle its logistics problems, the multiplier effect and donated resources are frequently used both by state Departments of Education and by local districts. The multiplier effect refers to the dissemination of information to a group whose members in their turn make a similar or prescribed dissemination to other groups. At the state level this may entail training "trainers" at regional levels who will then visit the local districts and individual target audiences as part of a statewide dissemination effort. At the local level, a presentation (with handouts or packets of background information) can be an effective way to reach the executive committee of an organization or the leaders of a targeted group. If the presentation is successful, these individuals can then make their own presentations or endorsements to their constituencies. In this way, it will be possible to contact a large number of people with little cost and effort. Such presentations by the leaders of a group or organization will further enhance the positive effects of dissemination. Donated resources may be in the form of free exposure by the media: newspapers, radio, television, community newsletters. The audience reached can be enormous; donated resources, therefore, are an important consideration for every dissemination plan.

Identifying Appropriate Media for Conveying Information

As evidenced in operating MCT programs, the medium and the format for carrying the information to an audience are important aspects of the dissemination plan. A key parameter in selecting or using a medium is the

amount and nature of the resources available (NSPRA, 1976); therefore, which media are selected by any given program will depend upon the particular circumstances within that program.

Available means for dissemination. In selecting the means for dissemination, planners may wish to consider the following two questions: (1) which means will reach each target audience most satisfactorily, and (2) what resources are available to support the means.

A brief discussion of the means of conveying information appears below. All are familiar; what may be unfamiliar to those inexperienced in a large-scale dissemination effort is the careful planning necessary to select the most appropriate means for each audience, so that the intended information is transmitted and the intended effect of that information is achieved. An inappropriate choice may be a waste of time, effort, and money, and may produce an adverse impact as well.

(1) CONFERENCES/WORKSHOPS. Generally these work best when the participation of the audience is desired for program development, implementation, or evaluation. Therefore, many programs use advisory or steering committees composed of local community members. Teachers come to workshops to learn how to develop the competency statements, prepare tests, or interpret and use test results. In South Burlington, Vermont, for example, a group of teachers attended a workshop during the summer of 1977 at which they developed assessments for the state-mandated competencies. Parents and other community groups may also be invited to help in reviewing the competencies or in setting the passing standards for the tests. However, to use the workshop or conference most effectively, it is often necessary to use other avenues for diffusing information in order to make audiences aware of the MCT program and to persuade them to participate.

(2) PUBLIC MEETINGS. The public forum can be very useful, providing the disseminator knows and understands fully the intent of the meeting and the composition of the attending audience before the presentation (NSPRA, 1976). For example, many public meetings relating to school or community affairs are attended by people who wish to participate in the decision making process. This participation may take the form of the community or group pressure which their presence can effect. In Oregon, for example,

the superintendent called for public meetings to be held statewide; the purpose of the meetings was to gauge public sentiment concerning what skills public schools should be responsible for teaching. A public meeting which has been called for the purpose of presenting information about the MCT program usually guarantees an interested audience. In Massachusetts public meetings are sponsored by the Department of Education in order to present and answer questions concerning the results of basic skills tests. In such meetings, the audience may consist largely of people who are very supportive of the school system and, it is to be hoped, of the program; at the other extreme, it may consist of people with negative attitudes toward the school system in general or toward the program itself. Thus, the disseminator should be prepared to cover the full gamut of possible reactions and queries in connection with the program. A second possibility for disseminating information at a public meeting is "piggybacking," or adding a dissemination effort to a meeting set for a different purpose: a school board meeting to discuss a controversial budget, or a town meeting to select and discuss political candidates. Many MCT program directors warn of the danger of inadvertently associating the program with other, perhaps emotionally laden issues and controversies on the regular agenda of a meeting. A meeting of people who have come to vote down a district budget, for example, may not be the appropriate occasion for disseminating information about a program which can itself arouse strong feelings.

(3) NEWSLETTERS/FLYERS/BROCHURES. Newsletters can be a very inexpensive means of informing the community of the MCT program and of keeping people informed of program progress. This is particularly true if a newsletter is the regular periodical of a school district, since the costs of adding the MCT program description will be relatively small. Nearly every MCT program uses this method. Flyers and brochures in a question-and-answer format have been found to be particularly useful. The Michigan Department of Education has produced several such flyers and brochures, as did the North Carolina Department of Education.

(4) NEWSPAPER ARTICLES. This medium generally reaches the largest number of people. Unfortunately, it can be the hardest to control in terms of accuracy and emphasis. Hubbell and Stech and the California Department of Education (1977) provide guidelines and approaches for achieving good media releases and interviews. Program planners may also want to consider Florida's use of

newspaper reporting in its program (Fisher, 1978). The Department permitted a small number of reporters to take the eleventh-grade Functional Literacy Test in order to foster greater understanding of the content and difficulty of the test among the public. In contacting the press, however, program personnel have found that material for publication may need to be reviewed carefully by the dissemination staff. Misinformation, once in print, may produce an effect which is difficult to overcome.

(5) TELEVISION BROADCASTS. Because of the high visual impact and wide exposure which this medium provides, it is important to present spokespersons who will appeal to particular audiences. For example, a good review of the MCT program by the spokesperson of a special interest group may be most beneficial in gaining that group's acceptance and support.

(6) RADIO BROADCASTS. There are generally two modes of radio presentation. One method is to have a newscaster present a capsule summary, perhaps periodically, of the MCT program. A consecutive set of presentations every day for a week may reach a diverse set of audiences. A second mode of dissemination might be in the form of a discussion or question-and-answer forum in which key school staff (or state-level staff) meet with a commentator or with the spokespersons of key target audiences to answer their questions and respond to their concerns. For example, Florida Department of Education staff members were interviewed by representatives of national radio networks to answer questions concerning the assessment program.

(7) WRITTEN SUMMARY REPORTS. Written reports are usually directed to a specialized audience, since interest will have to be at a fairly high level to ensure that the report will be read. However, reports take on added value as background information packets for use in multiplier-effect situations and with donated sources. In surveying summary reports for existing MCT programs, those programs with a record of success invariably have produced well-written reports in language which the general reader can understand. Such reports often emphasize program components and their rationales, topics which are known to be of interest and concern to a variety of public audiences. Michigan is one of a number of programs, for example, which prepares summary reports for various groups, including classroom teachers, parents, and school administrators.

(8) WRITTEN TECHNICAL REPORTS. This type of report is generally useful for state- or local-level planning, for local assessment of a program and the need for modification, and for assessing the value of each component of program planning, implementation, and evaluation. Reports prepared for current programs, such as the Michigan Educational Assessment Program, focus on important data presented in tables and charts so that a reader receives a comprehensive view of the program. Since such reports are generally available to the public, accuracy of content should be tightly monitored. These reports may also serve as resource material for media stories.

(9) MOVIES/SLIDE-TAPE SHOWS. Program planners may find these media to be more appropriate for disseminating information in a lively, topical fashion. In order to acquaint parents with the purpose of the testing program and the part they could play in strengthening their children's skills, the Michigan Department of Education prepared a filmstrip for district use. Slide/tape shows represent another avenue for conveying information, one that can be prepared in advance and used with many different groups.

It is frequently helpful to establish criteria for selecting the means for the diffusion of information. Some important issues which stand out in a review of existing programs are cost, available lead time, accessibility, and breadth of coverage and impact. Cost is always a prime consideration. Conferences, workshops, flyers, and brochures can be expensive due to high production and preparation costs and the size of such projects. Lead time may often mean that planning will take place weeks or even months before the information is expected to reach the intended audiences. For example, to mount their dissemination efforts, the Michigan Educational Assessment Program and the Massachusetts Basic Skills Assessment planned strategy and documents months before the programs were operative. Since a program may change or testing may occur before the time set for initiating the dissemination effort, it may be necessary to make readjustments in the dissemination plan.

The accessibility of the possible means for communication is another important consideration since scheduling depends on their availability, costs, and the work necessary for preparation or development (as in the case of documents). The breadth of coverage and the expected impact of a particular mode of dissemination are two factors to consider as well.

The form of the message for the vehicle of dissemination. Once the means has been selected, the content and format of information to be disseminated can be planned. If the method chosen is a workshop, then materials will be a consideration. Group presentations may require overhead transparencies, filmstrips, or handouts. The facts and summaries for the media may need to be carefully reviewed for accuracy, completeness, and impact not only as a whole, but in the light of the effect they will have as partial presentations. The choice of language for materials may often present problems. On the whole, it has been found that the avoidance of technical terms and concepts and of educational jargon is best.

The form of the message is also dependent on the circumstances in which the presentation of a message is to occur. It should certainly be useful to keep a record that will indicate what information is to be disseminated when, to whom, and how (Hubbell & Stech).

In preparing a plan for dissemination, it may also help to draw up charts with detailed descriptions in each box of the chart of the type of information to be disseminated, the vehicle, and the target audience(s). The use of planning charts will permit planners to map out the entire dissemination effort in outline form so that coordination, sequencing, and time commitments can be easily compared and grasped at a glance.

Documenting the Plan

It can be time-consuming to prepare a comprehensive plan for dissemination with stated procedures, and rationales for the suggestions and selections made. As with the planning of the form and content of dissemination presentations for particular audiences, the use of charts in other aspects of the planning process will permit the development of planning components in a clear and orderly manner. Charts also provide a systematic means of organizing a great deal of information in a format which is easy to understand and to explain to others. Two charts from state-level programs are presented here. The first is a timeline for the dissemination plan produced for one year of the Michigan Educational Assessment Program. The reader may notice that the dissemination tasks continue throughout the year including before, during, and after the test dates.

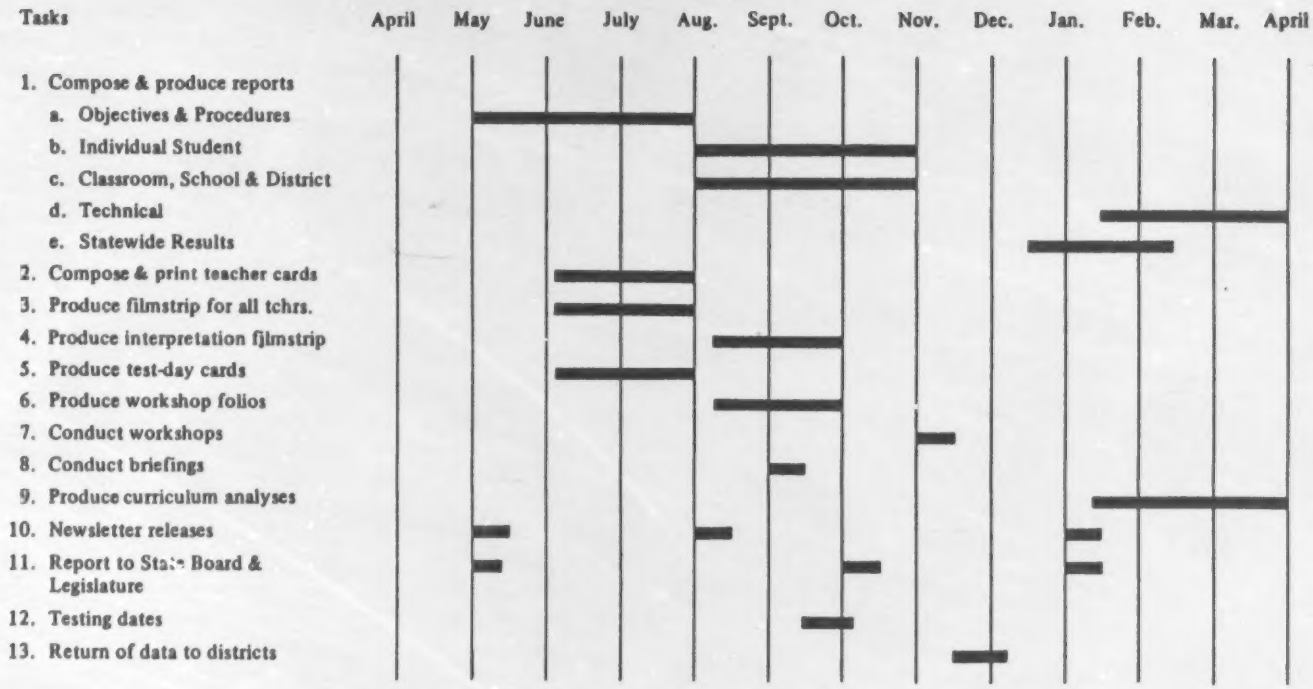
The second chart is from the California State Department of Education. It is a suggested means of producing an overall plan for managing the assessment information for local districts. The chart shows major audiences for dissemination and major sources of information.

Topic	Information to be reported (to)			Recordkeeping (by)		
	Student	Parent	Community and school board	Teachers	Principal's designee; e.g., counselor	Administrator
Standards of proficiency	1. Content of law 2. Skill areas to be covered 3. Proficiency levels	1. Content of law 2. Skill areas to be covered 3. Proficiency levels	1. Content of law 2. Skill areas to be covered 3. Proficiency levels	1. Monitoring of each student's progress in reaching required levels of proficiency	1. Standards to be covered in each department and course 2. Students who have and have not attained proficiency levels 3. Students on special projects	1. Criteria for proficiency assessment 2. Proficiencies to be included
Assessments	4. Frequency 5. Date and time 6. Individual results	4. Frequency 5. Date and time 6. Individual results	4. Frequency 5. Date and time 6. Group results	2. Methods of evaluation 3. Specific date and time 4. Results	4. Methods of assessment 5. Individual student results	3. Monitoring of all assessments 4. Statistical data for schools and districts
Conferences	7. Notification of conference 8. Status 9. Identification of alternatives	7. Letter announcing conference 8. Follow-up telephone call	7. Formative and summative data on conferences 8. Student and parent reaction	5. Participation in conferences	6. File copy of conference notifications 7. Follow-up phone calls 8. Date and time of conferences 9. File copies of decisions made at conferences, including alternative courses selected 10. Special projects	5. Verification or compliance 6. Provisions for alternatives
Instructional processes; alternatives	10. Courses available 11. Alternatives available	9. Courses available 10. Alternatives available	9. Courses available 10. Alternatives available	6. Standards to be covered in their course and department 7. Course alternatives	11. Standards to be covered in each course and department 12. Students on special projects 13. Alternatives to regular program	7. Courses of study and alternatives

* From Technical Assistance Guide for Proficiency Assessment, California, State Department of Education, 1977.

Figure II *

Timeline for Dissemination



* From Releasing Test Scores: Educational Assessment Program, How to Tell the Public, National School Public Relations Association, 1976.

Summary

In the case of a potentially controversial program, such as an MCT program, the dissemination effort may require more components and considerably more planning than that necessary for the report of an occurrence such as a sporting event in the local papers. The specter of accountability may be of concern to every identifiable audience: teachers, students, and administrators as well as parents, the news media, and special interest groups.

Dissemination, then, becomes a delicate and demanding set of activities ranging over the duration of the program. Consequently, it is important to recognize the need for comprehensive and careful planning in the early stages of an MCT program, so that dissemination activities can be fully integrated with the other elements of the program.

References

American Friends Service Committee. A citizen's introduction to minimum competency programs for students. Columbia, South Carolina: Southeastern Public Education Program, 1978.

California, State Department of Education. Technical assistance guide for proficiency assessment. Sacramento, California: Author, 1977.

Fisher, T. H. Florida's approach to competency testing. Phi Delta Kappan, 1978, 59(9), 599-602.

Hubbell, N. S. (Producer). Parent help at home pays off in school. Michigan Educational Assessment Program Filmstrip, 1978.

Hubbell, N. S., & Stech, E. L. Telling the testing story . . . through the mass media. Denver, Colorado: Colorado Department of Education, Cooperative Accountability Project, n.d.

Michigan, State Department of Education. Do YOU use MEAP test results appropriately? Lansing, Michigan: Author, n.d.

Michigan, State Department of Education. An educational health check. Lansing, Michigan: Author, n.d.

Michigan, State Department of Education. A pamphlet for parents. Lansing, Michigan: Author, n.d.

Michigan, State Department of Education. Questions and answers about the Michigan Educational Assessment Program. Lansing, Michigan: Author, n.d.

National School Public Relations Association (NSPRA). Releasing test scores: Educational assessment program, how to tell the public. Arlington, Virginia: Author, 1976.